HIGH-FIDELITY MONTE CARLO RECONSTRUCTION FROM FINITE-PHOTON SIMULATIONS WITH A PHOTON-BUDGET-CONDITIONED VOLUMETRIC STATE-SPACE MODEL

Anonymous authors
Paper under double-blind review

ABSTRACT

High-fidelity Monte Carlo (MC) simulation demands substantial sampling effort. We make the photon budget an explicit condition for learned reconstruction, allowing one volumetric model to adapt across eight sampling regimes while retaining MC as the forward solver. Photon-budget affine modulation coordinates directional spatial gating with Fourier-domain Mamba-3 processing. On a three-dimensional skin-lesion benchmark, the model improves mean PSNR by 4.82 dB over the strongest macro-average baseline (configuration-paired 95% interval: [4.57, 5.07] dB). Gains extend to inverse-transformed field summaries and tested held-out physical regimes, with capacity-matched controls supporting the design. Reconstruction from 10^3 photons exceeds raw 10^6-photon MC in mean PSNR, SSIM, and RMSE against matched references, using $1000\times$ fewer input photons. At 10^6 photons, MC plus inference achieves an $8.09\times$ same-GPU speedup over direct 10^8-photon MC, with a 0.67-dB mean PSNR gap against independent 10^9-photon references; a separate CPU-MC/GPU-inference workflow gives approximately $106\times$ timing speedup at the same budgets. These results establish photon-budget conditioning as a practical route to efficient volumetric MC reconstruction and motivate using known sampling effort as a design variable for learned correction in scientific simulation.

1 INTRODUCTION

Monte Carlo (MC) photon transport estimates light propagation, absorption, and energy deposition in heterogeneous scattering media (Wang et al., 1995; Fang & Boas, 2009; Zhu & Liu, 2013). Higher photon budgets reduce finite-sampling fluctuations but require more photon histories. Learning to reconstruct accurate three-dimensional (3D) fields from finite-photon simulations can therefore improve how sampling effort is used.

Finite-photon reconstruction uses adversarial learning, volumetric denoising, super-resolution, or Fourier processing, generally at specific budgets (Wang et al., 2022; Raayai Ardakani et al., 2022; Raayai-Ardakani et al., 2024; Fu et al., 2025). A unified model should exploit both cross-plane dependencies and the known sampling regime; state-space models provide suitable long-range representations (Gu & Dao, 2024; Liu et al., 2024; Zhu et al., 2024).

We introduce photon-budget-conditioned volumetric reconstruction: the known simulation budget becomes explicit context for a shared correction model. The model couples established spatial-frequency representations with photon-budget affine modulation to adapt volumetric reconstruction across eight budgets. This couples a controllable simulation resource to reconstruction without replacing the MC forward solver.

Configuration-paired statistics and inverse-transformed field metrics test the reliability and scope of gains; capacity-matched controls test the representation. Independent references, hardware-specific timings, and held-out physical regimes connect reconstruction to fidelity, cost, and extrapolation, supporting the joint choice of simulation effort and learned correction.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2 Related Work

2.1 Learning-Based Monte Carlo Acceleration

Learning-based MC acceleration uses reconstruction or direct surrogates. Reconstruction retains the solver and corrects finite-sampling outputs through adversarial, volumetric, or Fourier-domain models (Wang et al., 2022; Raayai Ardakani et al., 2022; Raayai-Ardakani et al., 2024; Fu et al., 2025). Direct surrogates such as NeuralRTE (Nguyen et al., 2025) instead predict transport responses from physical parameters. We retain MC and condition one volumetric reconstruction model on its known photon budget.

2.2 State-Space and Frequency-Domain Reconstruction

State-space models provide long-range representations for vision (Gu & Dao, 2024; Liu et al., 2024; Zhu et al., 2024), with Mamba-3 revising the underlying dynamics (Lahoti et al., 2026). FourierMamba (Li et al., 2025) models Fourier-space dependencies for image deraining, using zigzag scanning to organize spatial frequencies and a separate channel-frequency pathway. FMambaIR (Luan et al., 2025) combines spatial Mamba and Fourier branches for complementary degradation modeling in image restoration.

These works establish frequency-domain integration with Mamba. Our contribution is photon-budget-conditioned reconstruction of three-dimensional MC fields while retaining the forward solver. FACA processes a 3D complex spectrum with Mamba-3, channel-wise gain, real projection, and local modulation, jointly with directional gating and budget conditioning. This design targets correction across simulation budgets rather than introducing the general combination of Fourier processing and Mamba.

3 Method

3.1 Overall Architecture

Let $x, y \in \mathbb{R}^{1\times D\times H\times W}$ denote log-preprocessed finite-photon and matched 10^8-photon MC volumes, respectively, and let $n \in \{0, 1, \ldots, 7\}$ denote the photon-budget index. We use $\mathcal{A}$ and $\mathcal{A}^{-1}$ for the variance-stabilizing transform and its inverse, with $\widetilde{x} = \mathcal{A}(x)$ and $\widetilde{y} = \mathcal{A}(y)$.

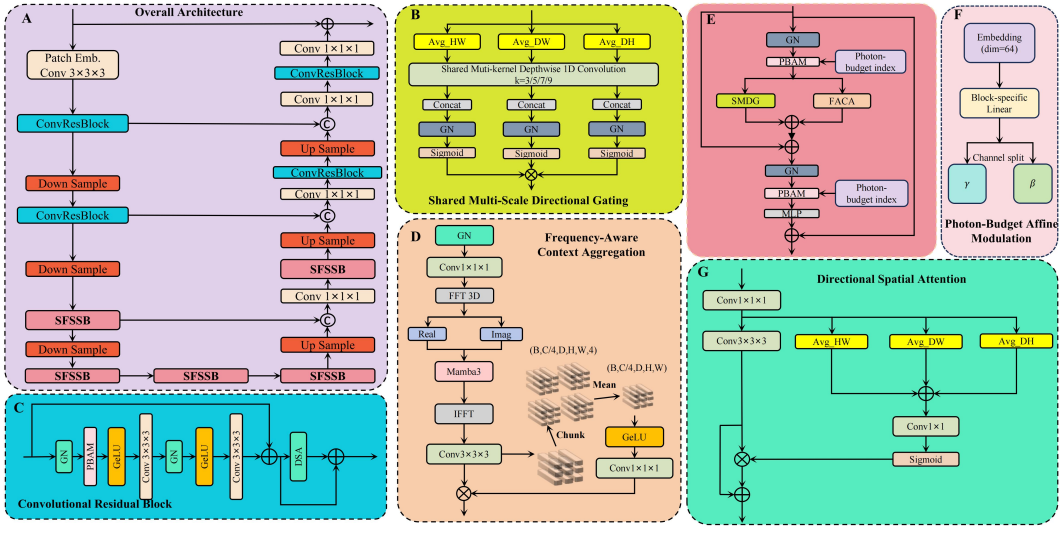

Figure 1: Proposed architecture: (A) hierarchical 3D encoder-decoder; (B) SMDG; (C) ConvResBlock; (D) FACA; (E) SFSSB; (F) photon-budget conditioning; (G) DSA.

As shown in Figure 1, the model is a hierarchical 3D encoder-decoder with channel widths 16→32→64→128. Convolutional Residual Blocks (ConvResBlocks) process the two highest-

resolution stages, while Spatial-Frequency State-Space Blocks (SFSSBs) operate at the lower-resolution stages and bottleneck. The decoder mirrors the hierarchy using transposed 3D convolutions and skip fusion. The three core components are Shared Multi-Scale Directional Gating (SMDG), Frequency-Aware Context Aggregation (FACA), and Photon-Budget Affine Modulation (PBAM); full architectural details are provided in Appendix A.

The network predicts a residual correction in the transformed domain:

$$\widehat{\widetilde{y}} = \widetilde{x} + \mathcal{R}_\theta(\widetilde{x}, n), \quad \widehat{y} = \mathcal{A}^{-1}(\widehat{\widetilde{y}}), \tag{1}$$

where $\mathcal{R}_\theta$ is the photon-budget-conditioned residual mapping. Thus, MC remains the physical forward solver and the network learns only the finite-sampling correction.

3.2 Spatial-Frequency State-Space Block

Each SFSSB combines directional spatial modeling and frequency-domain long-range modeling. MLP denotes two $1 \times 1 \times 1$ convolutions with an intermediate GELU. For block input X, group normalization (GN), and the block-specific PBAM operator Φ_b defined in Section 3.3,

$$\begin{aligned} U &= \Phi_b(\mathrm{GN}(X); n) \\ Z &= X + \mathrm{SMDG}(U) + \mathrm{FACA}(U) \\ X_{out} &= Z + \mathrm{MLP}(\Phi_b(\mathrm{GN}(Z); n)), \end{aligned} \tag{2}$$

SMDG summarizes U along the depth, height, and width axes. Each axis-preserving descriptor is divided into four channel groups and processed by shared depthwise 1D convolutions with kernel sizes 3, 5, 7 and 9. The resulting sigmoid gates A_D, A_H, A_W are broadcast to the volumetric feature:

$$\mathrm{SMDG}(U) = U \odot A_D \odot A_H \odot A_W, \tag{3}$$

where $\odot$ denotes element-wise multiplication with broadcasting. This provides multi-scale directional modulation while sharing filters across the three spatial axes.

FACA captures long-range context in the Fourier domain. A projected feature V undergoes a three-dimensional Fourier transform; its real and imaginary components are vectorized in native grid order as

$$Q = \mathrm{Flatten}([\mathrm{Re}(\mathcal{F}_{3D}(V)), \mathrm{Im}(\mathcal{F}_{3D}(V))]). \tag{4}$$

The resulting frequency sequence is processed by a Mamba-3 block (Lahoti et al., 2026), reassembled into a complex spectrum, modulated by a learnable channel-wise gain w_f, inverse transformed, and locally refined:

$$\begin{aligned} S = \mathrm{Conv}_{3\times3\times3}[\mathrm{Re}({\mathcal{F}_{3D}}^{-1}(\mathrm{complex}(\mathcal{M}_{M3}(Q)) \odot w_f))] \\ \mathrm{FACA}(U) = S \odot G. \end{aligned} \tag{5}$$

Here, $\mathcal{M}_{M3}$ denotes Mamba-3 and G is obtained by grouping and aggregating the channels of S. Exact operator configurations are given in Appendix A.

FACA explicitly takes the real part after inverse transformation; the learned spectrum need not retain Hermitian symmetry (Appendix A.5).

3.3 Photon-Budget Conditioning

The eight photon budgets, $10^2, 10^3, 10^4, 5\times10^4, 10^5, 2\times10^5, 5\times10^5, 10^6$, are indexed by n and explicitly supplied to the reconstruction model. An embedding table maps n to $c_n = E[n] \in \mathbb{R}^{64}$. For block b with C_b channels, a block-specific projection produces $\gamma_b, \beta_b \in \mathbb{R}^{C_b}$:

$$[\gamma_b,\ \beta_b] = P_b(c_n), \quad \Phi_b(F; n) = F \odot (1 + \gamma_b) + \beta_b, \tag{6}$$

The projection is zero-initialized, so $\Phi_b(F; n) = F$ at initialization. PBAM therefore allows one shared backbone to adapt its intermediate representations to the known finite-sampling regime without changing the reconstruction checkpoint.

3.4 Training Objective

For MC absorption values a, preprocessing computes $u = \ln(1+e^8 a)$ and resamples to 64^3, yielding x, y. We apply the Anscombe transform (Anscombe, 1948) to these stored volumes:

$$\mathcal{A}(x) = 2\sqrt{\max(x, 0) + \frac{3}{8}}, \quad \mathcal{A}^{-1}(z) = \max[(\frac{z}{2})^2 - \frac{3}{8}, 0]. \tag{7}$$

Training combines four terms. $\mathcal{L}_{wMSE}$ is a photon-weighted transformed-domain mean squared error (MSE); $\mathcal{L}_{SSIM}$ is a three-dimensional extension of the structural similarity index measure (SSIM) loss (Wang et al., 2004), with a mini-batch-dependent coefficient that increases when low-photon samples are more prevalent. For photon budgets $\geq 2 \times 10^5$, $\mathcal{L}_{raw}$ adds log-domain fidelity after inverse Anscombe, while at 10^6 photons $\mathcal{L}_{id}$ discourages unnecessary deviation from the comparatively stable finite-photon input. The complete objective is

$$\mathcal{L} = \mathcal{L}_{wMSE} + \lambda_s\,(\mathcal{B})\,\mathcal{L}_{\text{SSIM}} + 0.08\,\mathcal{L}_{raw} + 0.03\,\mathcal{L}_{id}. \tag{8}$$

The final two terms are evaluated only for eligible samples. Full loss definitions, architectural hyperparameters, and optimization details are provided in Appendices A–C.

4 Experiments

4.1 Experimental Setup

We use MCmatlab (Marti et al., 2018) to simulate multilayer Asian-skin models with an irregular basal cell carcinoma (BCC)-like lesion, at 633 nm under normally incident pencil-beam illumination. A six-factor sweep covers melanin, dermal, and lesion absorption, lesion reduced scattering, lesion scale, and invasion depth. We select 405 of 3,240 candidate configurations; Appendix D details the benchmark.

Each configuration is simulated at eight budgets and paired with a matched 10^8-photon reference. A configuration-level 243/81/81 training/validation/test split is propagated across budgets to prevent leakage. PSNR, SSIM, and RMSE use complete log-domain volumes after inverse Anscombe. Paired uncertainty and inverse-transformed field errors supplement these metrics. Independent 10^9-photon references evaluate the same 81 tests without retraining or checkpoint selection. Timing uses one fixed configuration; held-out experiments use separate splits and retrained models. Appendices C–F give the protocols.

Baselines include 3D U-Net and DnCNN-U-Net (Ronneberger et al., 2015; Çiçek et al., 2016; Zhang et al., 2017; Raayai Ardakani et al., 2022), 3D FNO (Kovachki et al., 2023), MC-Fourier (Fu et al., 2025), ReconFormer (Guo et al., 2024), VEMamba (Gao & Gao, 2026), and DenoMamba (Öztürk et al., 2026); source implementations and task adaptations are given in Appendix D.

4.2 Reconstruction Across Photon Budgets

Table 1 compares reconstruction fidelity at 10^3, 10^5, and 10^6 photons. The proposed model achieves the highest PSNR and lowest RMSE at all three representative budgets; its SSIM is highest at 10^3, tied highest at 10^6, and slightly below 3D FNO at 10^5 (0.9994 versus 0.9996). Complete eight-budget PSNR, SSIM, and RMSE results are reported in Appendix Tables E1–E3.

Table 1: Reconstruction fidelity at representative photon budgets.

Method	10^3 photons			10^5 photons			10^6 photons		
	PSNR↑	**SSIM↑**	**RMSE↓**	**PSNR↑**	**SSIM↑**	**RMSE↓**	**PSNR↑**	**SSIM↑**	**RMSE↓**
Raw MC	8.9005	0.0294	5.0193	23.8414	0.7163	0.9053	38.6475	0.9597	0.1643
3D U-Net	32.8294	0.8912	0.3218	39.6803	0.9756	0.1452	45.0474	0.9934	0.0784
DnCNN-U-Net	33.8349	0.8988	0.2864	44.0514	0.9916	0.0878	49.3451	0.9985	0.0479
3D FNO	37.2605	0.9947	0.2176	48.4199	**0.9996**	0.0587	49.4954	**0.9997**	0.0563
MC-Fourier	37.7187	0.9890	0.1856	46.0165	0.9972	0.0706	48.4882	0.9981	0.0540
ReconFormer	35.8789	0.9450	0.2277	48.6355	0.9966	0.0519	52.6980	0.9981	0.0325
VEMamba	39.8138	0.9885	0.1451	46.8049	0.9970	0.0642	50.4737	0.9987	0.0424
DenoMamba	35.8695	0.9680	0.2282	47.6949	0.9971	0.0580	51.7654	0.9991	0.0366
Proposed	**42.6318**	**0.9963**	**0.1078**	**53.1098**	0.9994	**0.0316**	**56.7556**	**0.9997**	**0.0206**

Reconstruction improves sampling efficiency: at 10^3 photons, it reaches 42.6318 dB PSNR, 0.9963 SSIM, and 0.1078 RMSE, exceeding raw 10^6-photon MC (38.6475 dB, 0.9597, 0.1643) against the same matched 10^8 reference. This uses 1000 times fewer input photon histories at better mean log-domain metrics, not a $1000\times$ runtime speedup. Figure 2 illustrates error reduction and preserved depth-dependent structure. The proposed model leads the architectural baselines in PSNR/RMSE from 10^3 to 10^6; at 10^2, 3D FNO has higher PSNR and lower RMSE, and VEMamba also has lower RMSE, while the proposed model has the highest SSIM.

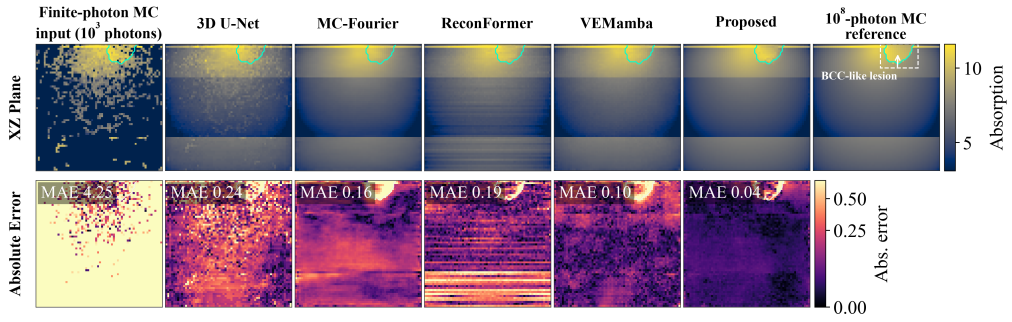

Figure 2: Qualitative reconstruction comparison at 10^3 photons. Top: reconstructed XZ planes. Bottom: absolute error relative to the matched 10^8-photon MC reference.

Configuration-paired analysis strengthens this comparison: the eight-budget mean gain over ReconFormer, the strongest macro-average baseline, is 4.82 dB (95% interval [4.57, 5.07]), with positive configuration-level gains for all 81 tests. Intervals against the strongest baseline at each budget are above zero at seven budgets; 10^2 remains inconclusive (Table E10). After inverse log transformation, mean integrated absorption and depth-profile errors are 3.96% and 7.76%, respectively, 20.35% and 17.43% lower than MC-Fourier (Table E11). Thus, gains extend to integral and depth-dependent field summaries on the evaluation grid.

4.3 Slice-Wise versus Volumetric Reconstruction

Table 2 compares the proposed model with an architecture- and split-matched slice-wise counterpart. At 10^3 photons, volumetric reconstruction improves PSNR by 4.1243 dB and reduces RMSE by 36.0%; the advantage persists across all eight photon budgets (Appendix Table E4).

Figure 3 provides the corresponding spatial evidence. Independent XZ-slice reconstruction introduces stripe artifacts and reduced consistency in the unprocessed XY and YZ views, whereas volumetric reconstruction maintains coherent structure across all three planes. These results directly support cross-plane context as a useful reconstruction signal under finite-photon sampling.

Table 2: Slice-wise versus volumetric reconstruction fidelity at representative photon budgets.

Model	10^3 photons			10^5 photons			10^6 photons		
	PSNR↑	**SSIM↑**	**RMSE↓**	**PSNR↑**	**SSIM↑**	**RMSE↓**	**PSNR↑**	**SSIM↑**	**RMSE↓**
Slice-wise	38.5075	0.9782	0.1685	50.2154	0.9989	0.0437	54.9655	0.9995	0.0251
Proposed	**42.6318**	**0.9963**	**0.1078**	**53.1098**	**0.9994**	**0.0316**	**56.7556**	**0.9997**	**0.0206**

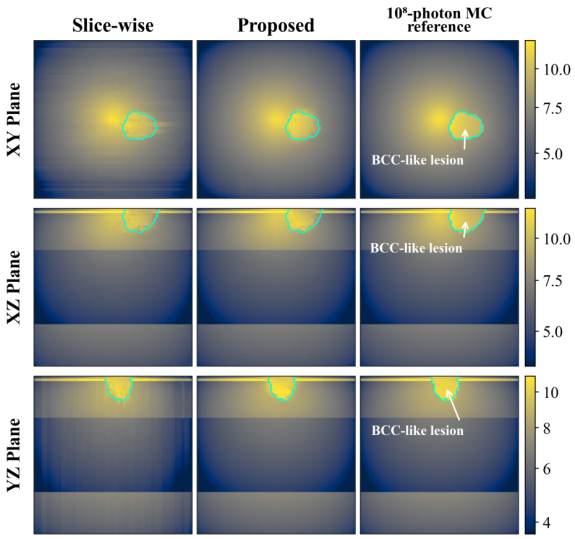

Figure 3: Orthogonal-plane reconstruction at 10^3 photons for the slice-wise counterpart, proposed model, and 10^8-photon MC reference.

4.4 Unified Photon-Budget-Conditioned Reconstruction

Table 3 compares eight separately trained photon-specific models, a unified model without photon-budget conditioning, and the proposed model under matched held-out configurations and training-pair counts per model. The proposed model achieves the highest average PSNR and SSIM while replacing eight photon-specific networks totaling 33.0374 M parameters with one 4.2092 M-parameter network, reducing the aggregate parameter footprint by approximately 7.8×. Its average RMSE (0.0780) remains slightly above the photon-specific value (0.0762).

Table 3: Photon-specific and unified reconstruction across eight photon budgets.

Strategy	Photon-budget conditioning	No. of models	Avg. PSNR↑	Avg. SSIM↑	Avg. RMSE↓	Total params (M)
Photon-specific	—	8	49.1926	0.9975	**0.0762**	33.0374
Unified, no conditioning	×	1	48.4284	0.9971	0.0826	4.1297
Proposed	✓	1	**49.5687**	**0.9976**	0.0780	4.2092

Conditioning itself contributes substantially: removing photon-budget conditioning reduces average PSNR from 49.5687 to 48.4284 dB and increases RMSE from 0.0780 to 0.0826. Complete photon-specific results are given in Appendix Table E5.

Appendix Figure E1 and Tables E5–E6 show regime-dependent benefits: performance is mixed at 10^2–10^4 photons; from 5×10^4 to 10^6, the proposed model has the highest PSNR and lowest RMSE, and highest or tied-highest SSIM among these three strategies.

4.5 Ablation and Mechanistic Analysis

Table 4 retains high-resolution DSA in all seven variants. SMDG improves Base PSNR by 1.1778 dB. Capacity-matched convolutional replacements differ from FACA variants by only 2,790 parameters; FACA improves PSNR by 1.3234 dB without SMDG and 1.6732 dB with it. A capacity-matched spatial-domain Mamba control reaches 48.0089 dB versus 48.4284 dB for FACA, which has higher observed means at all budgets (Table E12). PBAM further improves F to the final model. These controls support FACA relative to the tested alternatives; Appendix E.4 details the component comparisons.

Table 4: Component and capacity-controlled ablation, macro-averaged across eight photon budgets. All variants retain the same high-resolution DSA blocks.

Variant	SMDG	FACA	PBAM	PSNR ↑	SSIM ↑	RMSE ↓	Params (M)	FLOPs† (G)
A. Base	×	×	×	45.9624	0.9921	0.1041	1.064	37.12
B. + SMDG	✓	×	×	47.1402	0.9969	0.0899	1.071	37.12
C. Matched Conv	×	×	×	46.3127	0.9964	0.0986	4.120	43.42
D. + FACA	×	✓	×	47.6361	0.9969	0.0897	4.123	43.40
E. SMDG + Conv	✓	×	×	46.7552	0.9965	0.0945	4.127	43.42
F. SMDG + FACA	✓	✓	×	48.4284	0.9971	0.0826	4.130	43.40
G. Proposed	✓	✓	✓	**49.5687**	**0.9976**	**0.0780**	4.209	43.40

C/E use a capacity-matched Conv control in place of FACA. †Counted with `torch.utils.flop_counter` for a $1 \times 1 \times 64^3$ input; unsupported FFT/IFFT and selective-scan operations are excluded.

Figure 4 shows partially distinct error reductions from SMDG and FACA, with their combination extending the corrected regions. Together with the interventions and capacity-matched controls, this supports complementary contributions within the evaluated model.

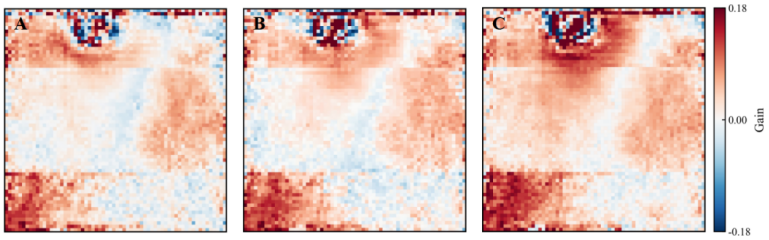

Figure 4: Spatial error reduction from SMDG, FACA, and their combination at 10^4 photons.

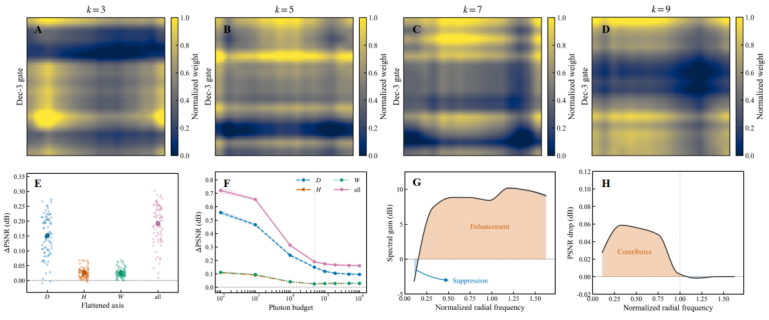

Figure 5: Mechanistic analysis of SMDG and FACA: directional gates and gate-flattening sensitivity (A–F), radial spectral gain (G), and frequency-shell intervention (H).

The component contributions are budget dependent: SMDG provides the larger individual PSNR gain at 10^2–10^3, whereas FACA contributes more from 10^4 to 10^6; photon-budget conditioning

becomes consistently beneficial from 5×10^4 onward. Figure 5 tests whether these learned representations are functionally used. Mean-preserving flattening of SMDG gates decreases PSNR across all budgets, while reverting low-to-mid-frequency shells to their pre-FACA representations reduces PSNR, supporting functional dependence on both directional gating and modified frequency content.

4.6 Independent-Reference Fidelity and Runtime

Table 5 evaluates unchanged reconstructions against matched 10^8-photon and independently simulated 10^9-photon references. Each PSNR uses its own reference's dynamic range, so the columns are not interchangeable. Against the common 10^9 reference, 10^6-photon reconstruction achieves 54.8538 dB versus 55.5246 dB for direct 10^8-photon MC, a 0.6708-dB difference in mean PSNR.

Table 5: Reconstruction fidelity and same-GPU runtime across photon budgets.

Photon budget	PSNR vs. 10^8	PSNR vs. 10^9	GPU-MC time (s)	MC + inference time (s)	Speedup
10^2	34.0070	34.9975	0.7700	0.7926	8.26×
10^3	42.6318	43.4434	0.7501	0.7726	8.48×
10^4	47.7852	48.2478	0.7423	0.7649	8.56×
5×10^4	51.4834	51.2789	0.7694	0.7919	8.27×
10^5	53.1098	52.5179	0.7786	0.8012	8.17×
2×10^5	54.5029	53.4508	0.7832	0.8057	8.13×
5×10^5	56.2741	54.4979	0.7645	0.7870	8.32×
10^6	56.7556	**54.8538**	0.7870	0.8095	8.09×
Direct 10^8 MC	—	55.5246	6.5482	6.5482	1×

PSNR (dB) is averaged over 81 test volumes. Both timing components use the same RTX 5080; MC timing uses one fixed configuration. Speedup is relative to direct 10^8-photon GPU-MC. The sum excludes transforms, data transfer, and I/O. Runtime SDs and the full protocol appear in Appendix E.5.

The $8.09\times$ same-GPU speedup carries a 0.6708-dB gap. At the same 10^6-to-10^8 budgets, CPU-MC plus GPU inference takes 11.4930 s versus 1213.6854 s for direct CPU-MC, approximately $106\times$ faster (Table E9). This separate heterogeneous workflow shares the timing exclusions above. Neither ratio establishes equal-quality acceleration. For the $1000\times$ photon-saving comparison, 10^3-photon MC plus inference takes 0.7726 s versus 0.7870 s for raw 10^6-photon GPU-MC, only $1.02\times$ faster. Reconstructed-budget GPU runtimes span 0.7649–0.8095 s, so photon savings do not translate proportionally into latency savings.

4.7 Held-Out Physical-Regime Generalization

We train each method from scratch on two separate splits of the 405 configurations. Lesion absorption trains on $\mu_{a,\text{lesion}} \in \{0.5, 1.0\}$ cm^{-1} and holds out 1.5; lesion scale trains on FWHM $\{0.5, 0.8, 1.2\}$ mm and holds out 1.6. Disjoint, unpaired ID/held-out sets contain 39/142 and 46/101 configurations, respectively, with ID values nearest the held-out regime. Other parameter distributions are approximately matched. Held-out data enter neither training nor model selection (Appendix F).

The proposed model has the highest held-out PSNR in both tests (Table 6). Its absorption drop is 0.14 dB; the larger lesion-scale drop is 1.99 dB, while held-out PSNR exceeds MC-Fourier and 3D FNO by 2.31 and 3.87 dB. MC-Fourier has a smaller scale drop (1.84 dB), so highest held-out quality does not imply least degradation. These tests cover the stated absorption and scale values, not arbitrary shapes; their different splits also preclude ranking shift difficulty from absolute PSNR.

Table 6: Held-out physical-regime generalization. PSNR (dB) is macro-averaged across eight photon budgets in the log domain against matched 10^8-photon references. Drop is ID minus held-out PSNR.

	Lesion absorption			Lesion scale		
Method	ID PSNR	Held-out PSNR	Drop	ID PSNR	Held-out PSNR	Drop
3D FNO	42.52	41.80	0.72	41.74	39.19	2.55
MC-Fourier	42.82	42.58	0.24	42.59	40.75	1.84
Proposed	48.07	47.93	0.14	45.05	43.06	1.99

Configuration-bootstrap 95% drop intervals are [-0.75, 1.00] dB for absorption and [1.44, 2.53] dB for lesion scale. Thus, the absorption-drop estimate is compatible with zero, whereas scale extrapolation shows a clearer loss. Full metrics and uncertainty conditional on each trained model appear in Appendix F.

5 Discussion and Limitations

The key result is that known simulation effort improves learned reconstruction when supplied as an explicit model input. Photon-budget conditioning improves mean PSNR by 1.14 dB; one model also replaces eight photon-specific networks with a $7.8\times$ smaller aggregate parameter footprint and higher mean PSNR and SSIM, although slightly higher RMSE. Together with paired comparisons, inverse-transformed field metrics, and capacity-matched controls, these results support a unified correction strategy across sampling regimes. Building on existing frequency-domain Mamba work, the contribution connects budget adaptation to volumetric MC reconstruction while retaining the forward solver.

This connection matters for how computation is allocated. The measured operating points offer $1000\times$ fewer input photons to exceed a raw-MC quality level, or $8.09\times$ same-GPU speedup with a 0.67-dB independent-reference gap; the separate CPU-MC/GPU-inference workflow gives approximately $106\times$ timing speedup at the latter budgets. These distinct gains support selecting simulation effort jointly with learned correction, expanding practical quality–cost choices without replacing MC. They are not equal-quality timing guarantees and exclude offline costs. The broader opportunity is to condition learned correction on known sampling effort in other stochastic simulators, subject to validation of their error structure and fidelity requirements.

Evidence covers one 633-nm skin-lesion benchmark, source configuration, 64^3 grid, and eight budgets. Held-out tests address specified absorption and scale shifts; other physics, arbitrary shapes, and unseen budgets remain untested. Bootstrap intervals exclude training variability. Inverse-transformed summaries establish neither native-grid accuracy nor conservation or downstream performance. FACA's real projection does not assume negligible imaginary features. References retain sampling error; timing covers one configuration and excludes transforms, transfers, and I/O.

6 Conclusion

We establish photon-budget conditioning as an effective way to connect sampling effort with learned volumetric MC reconstruction. A shared spatial-frequency model delivers strong reconstruction quality across eight budgets while retaining the MC solver, supported by paired statistics, field-level evaluation, capacity-matched controls, and targeted extrapolation tests. The measured photon savings and GPU and CPU workflow speedups translate reconstruction gains into a wider range of computational costs and fidelity levels for this simulation workflow. More broadly, the results motivate using known sampling effort as a design variable for learned correction in scientific computing.

AI USE STATEMENT

Generative AI was used to assist with LaTeX conversion, reference formatting, language editing, and conference-template typesetting. It also assisted with restructuring the presentation of author-provided contributions and results, discussing their interpretation and scope, checking consistency among numerical values, equations, and code, and proposing and implementing code corrections, tests, and documentation edits. The authors are responsible for reviewing and verifying all AI-assisted material and take full responsibility for the final content of the paper and accompanying code.

REFERENCES

F. J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3–4):246–254, 1948. doi: 10.1093/biomet/35.3-4.246.

A. N. Bashkatov, E. A. Genina, V. I. Kochubey, and V. V. Tuchin. Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000 nm. *Journal of Physics D: Applied Physics*, 38(15):2543–2555, 2005. doi: 10.1088/0022-3727/38/15/004.

Alexey N. Bashkatov, Elina A. Genina, and Valery V. Tuchin. Optical properties of skin, subcutaneous, and muscle tissues: a review. *Journal of Innovative Optical Health Sciences*, 4(1):9–38, 2011. doi: 10.1142/S1793545811001319.

Frank P. Bolin, Luther E. Preuss, Roy C. Taylor, and Robert J. Ference. Refractive index of some mammalian tissues using a fiber optic cladding method. *Applied Optics*, 28(12):2297–2303, 1989. doi: 10.1364/AO.28.002297.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 424–432, 2016. doi: 10.1007/978-3-319-46723-8_49.

Qianqian Fang and David A. Boas. Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units. *Optics Express*, 17(22):20178–20190, 2009. doi: 10.1364/OE.17.020178.

Shaoduo Fu, Yipin Lv, Pengwei Han, Jiayue Wang, Jiacheng Man, Bingxue Zhang, Kun Wang, Zhanxiu Cai, Feng Gao, and Jiao Li. High-speed computation of high-quality light fluence distribution from low-photon Monte Carlo using a Fourier neural network. *Optics Letters*, 50(17):5302–5305, 2025. doi: 10.1364/OL.566784.

Longmi Gao and Pan Gao. VEMamba: Efficient isotropic reconstruction of volume electron microscopy with axial-lateral consistent Mamba. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026. URL https://github.com/I2-Multimedia-Lab/VEMamba.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *First Conference on Language Modeling (COLM)*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Pengfei Guo, Yiqun Mei, Jinyuan Zhou, Shanshan Jiang, and Vishal M. Patel. ReconFormer: Accelerated MRI reconstruction using recurrent Transformer. *IEEE Transactions on Medical Imaging*, 43(1):582–593, 2024. doi: 10.1109/TMI.2023.3314747.

Nikola B. Kovachki, Zong-Yi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://www.jmlr.org/papers/v24/21-1524.html.

Aakash Lahoti, Kevin Y. Li, Berlin Chen, Caitlin Wang, Aviv Bick, J. Zico Kolter, Tri Dao, and Albert Gu. Mamba-3: Improved sequence modeling using state space principles. *International Conference on Learning Representations (ICLR)*, 2026. URL https://openreview.net/forum?id=HwCvaJOiCj.

Dong Li, Yidi Liu, Xueyang Fu, Jie Huang, Senyan Xu, Qi Zhu, and Zheng-Jun Zha. FourierMamba: Fourier learning integration with state space models for image deraining. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 35342–35365. PMLR, 2025. URL https://proceedings.mlr.press/v267/li25be.html.

Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. VMamba: Visual state space model. *Advances in Neural Information Processing Systems*, 37:103031–103063, 2024.

Xin Luan, Huijie Fan, Qiang Wang, Nan Yang, Shiben Liu, Xiaofeng Li, and Yandong Tang. FMambaIR: A hybrid state-space model and frequency domain for image restoration. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–14, 2025. Article 4201614. doi: 10.1109/TGRS.2025.3526927.

Dominik Marti, Rikke N. Aasbjerg, Peter E. Andersen, and Anders K. Hansen. MCmatlab: An open-source, user-friendly, MATLAB-integrated three-dimensional Monte Carlo light transport solver with heat diffusion and tissue damage. *Journal of Biomedical Optics*, 23(12):121622, 2018. doi: 10.1117/1.JBO.23.12.121622.

Igor V. Meglinski and Stephen J. Matcher. Quantitative assessment of skin layers absorption and skin reflectance spectra simulation in the visible and near-infrared spectral regions. *Physiological Measurement*, 23(4):741–753, 2002. doi: 10.1088/0967-3334/23/4/312.

Vinh Nguyen, Abigail Clennell, Vladislav V. Yakovlev, and Alexander Doronin. NeuralRTE: Forward photon transport simulations in turbid media aided by machine learning and artificial intelligence. *Computers in Biology and Medicine*, 196:110892, 2025. doi: 10.1016/j.compbiomed.2025.110892.

Şaban Öztürk, Oğuz Can Duran, and Tolga Çukur. DenoMamba: A fused state-space model for low-dose CT denoising. *IEEE Journal of Biomedical and Health Informatics*, 30(6):5353–5366, 2026. doi: 10.1109/JBHI.2025.3629034.

John H. Pyne, Esther Myint, Elizabeth M. Barr, Simon P. Clark, and Ruihang Hou. Basal cell carcinoma: Variation in invasion depth by subtype, sex, and anatomic site in 4,565 cases. *Dermatology Practical & Conceptual*, 8(4):314–319, 2018. doi: 10.5826/dpc.0804a13.

Matin Raayai Ardakani, Leiming Yu, David R. Kaeli, and Qianqian Fang. Framework for denoising Monte Carlo photon transport simulations using deep learning. *Journal of Biomedical Optics*, 27(8):083019, 2022. doi: 10.1117/1.JBO.27.8.083019.

Matin Raayai-Ardakani, David R. Kaeli, and Qianqian Fang. Accelerating 3-D Monte Carlo photon transport simulations by leveraging deep-learning based volume super-resolution methods. *Optica Biophotonics Congress: Biomedical Optics 2024*, paper OS3D.7, 2024. doi: 10.1364/OTS.2024.OS3D.7.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015. doi: 10.1007/978-3-319-24574-4_28.

R. B. Saager, D. J. Cuccia, S. Saggese, K. M. Kelly, and A. J. Durkin. A light emitting diode (LED) based spatial frequency domain imaging system for optimization of photodynamic therapy of nonmelanoma skin cancer: Quantitative reflectance imaging. *Lasers in Surgery and Medicine*, 45(4):207–215, 2013. doi: 10.1002/lsm.22139.

Elena Salomatina, Brian Jiang, John Novak, and Anna N. Yaroslavsky. Optical properties of normal and cancerous human skin in the visible and near-infrared spectral range. *Journal of Biomedical Optics*, 11(6):064026, 2006. doi: 10.1117/1.2398928.

Kerry Setchfield, Alistair Gorman, A. Hamish R. W. Simpson, Michael G. Somekh, and Amanda J. Wright. Effect of skin color on optical properties and the implications for medical optical technologies: A review. *Journal of Biomedical Optics*, 29(1):010901, 2024. doi: 10.1117/1.JBO.29.1.010901.

Yu Shimojo, Takahiro Nishimura, Hisanao Hazama, Toshiyuki Ozawa, and Kunio Awazu. Measurement of absorption and reduced scattering coefficients in Asian human epidermis, dermis, and subcutaneous fat tissues in the 400- to 1100-nm wavelength range for optical penetration depth and energy deposition analysis. *Journal of Biomedical Optics*, 25(4):045002, 2020. doi: 10.1117/1.JBO.25.4.045002.

Lihong Wang, Steven L. Jacques, and Liqiong Zheng. MCML—Monte Carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, 47(2):131–146, 1995. doi: 10.1016/0169-2607(95)01640-F.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Shiyuan Wang, Jiaming Chen, Fengdi Zhang, Mingyang Zhao, Xiaoyu Cui, and Shuo Chen. Accelerating Monte Carlo simulation of light propagation in tissue mimicking turbid medium based on generative adversarial networks. *Medical Physics*, 49(2):1209–1215, 2022. doi: 10.1002/mp.15350.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. doi: 10.1109/TIP.2017.2662206.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. *European Conference on Computer Vision (ECCV 2018)*, pp. 294–310, 2018. doi: 10.1007/978-3-030-01234-2_18.

Caigang Zhu and Quan Liu. Review of Monte Carlo modeling of light transport in tissues. *Journal of Biomedical Optics*, 18(5):050902, 2013. doi: 10.1117/1.JBO.18.5.050902.

Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision Mamba: Efficient visual representation learning with bidirectional state space model. *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 62429–62442, 2024. URL https://proceedings.mlr.press/v235/zhu24f.html.

A Detailed Architecture and Training Configuration

A.1 Network Configuration

Module abbreviations and symbols retain their definitions from the main text; only additional symbols are introduced below. The proposed volumetric model operates on single-channel 3D Monte Carlo (MC) volumes of spatial size

$$64 \times 64 \times 64$$

The encoder-decoder contains four feature scales with channel dimensions

$$16,\ 32,\ 64,\ 128$$

corresponding to spatial resolutions

$$64^3,\ 32^3,\ 16^3,\ 8^3$$

respectively.

The detailed network configuration is summarized in Table A1.

Table A1: Detailed configuration of the proposed volumetric model.

Stage	**Resolution**	**Channels**	**Operation**
Input	64^3	1	Anscombe-transformed log-preprocessed MC volume
Embedding	64^3	16	$3 \times 3 \times 3$ Conv
Encoder 1	64^3	16	ConvResBlock
Downsample 1	32^3	32	$3 \times 3 \times 3$ Conv, stride 2
Encoder 2	32^3	32	ConvResBlock
Downsample 2	16^3	64	$3 \times 3 \times 3$ Conv, stride 2
Encoder 3	16^3	64	SFSSB
Downsample 3	8^3	128	$3 \times 3 \times 3$ Conv, stride 2
Bottleneck	8^3	128	3 SFSSB
Decoder 3	16^3	64	Up-conv + skip fusion + SFSSB
Decoder 2	32^3	32	Up-conv + skip fusion + ConvResBlock
Decoder 1	64^3	16	Up-conv + skip fusion + ConvResBlock
Output	64^3	1	$1 \times 1 \times 1$ Conv + global residual

Each downsampling layer consists of a stride-2 $3 \times 3 \times 3$ convolution followed by Group Normalization and GELU. Upsampling is implemented using a $2 \times 2 \times 2$ transposed convolution with stride 2. At each decoder stage, the upsampled feature is concatenated with the encoder feature at the corresponding resolution and fused using a $1 \times 1 \times 1$ convolution.

The network predicts a residual correction in the variance-stabilized domain:

$$\widehat{\widetilde{y}} = \widetilde{x} + \mathcal{R}_\theta\left(\widetilde{x}, n\right) \tag{A.1}$$

Equation (A.1) uses the transformed-domain notation of Section 3.1 and the global residual connection after the final $1 \times 1 \times 1$ projection.

A.2 ConvResBlock and DSA

ConvResBlock is used at the 64^3 and 32^3 stages to emphasize local volumetric representation while avoiding long state-space sequences at high spatial resolution.

For the input feature X, H_1 and H_2 denote the outputs of the first and second convolutional transformations, respectively. The first transformation is

$$H_1 = \mathrm{Conv}_{3\times3\times3}\left[\mathrm{GELU}\left(\Phi_b\ \left(\mathrm{GN}_1(X); n\right)\right)\right] \tag{A.2}$$

GN_1 and GN_2 denote the two normalization layers in ConvResBlock; Φ_b and c_n follow Section 3.3.

The second transformation is

$$H_2 = \mathrm{Conv}_{3\times3\times3}\left[\mathrm{GELU}\left(\mathrm{GN}_2\left(H_1\right)\right)\right] \tag{A.3}$$

The convolutional residual output X_r is

$$X_r = X + H_2 \tag{A.4}$$

Directional Spatial Attention (DSA) is subsequently applied through a second residual connection:

$$X_{out} = X_r + \mathrm{DSA}\left(X_r\right) \tag{A.5}$$

DSA first projects its input using a $1 \times 1 \times 1$ convolution. Three directional descriptors, P_D, P_H, and P_W, are then generated by retaining one spatial axis while averaging over the other two:

Let X' denote this projected feature.

$$P_D = \mathrm{Avg}_{H,W}(X'),\ P_H = \mathrm{Avg}_{D,W}(X'),\ P_W = \mathrm{Avg}_{D,H}(X') \tag{A.6}$$

The descriptors are summed, projected, and passed through the sigmoid function $\sigma(\cdot)$ to obtain the attention map A_{DSA}:

$$A_{\mathrm{DSA}} = \sigma\left[\mathrm{Conv}_{1\times1\times1}\left(P_D + P_H + P_W\right)\right] \tag{A.7}$$

In parallel, a $3 \times 3 \times 3$ convolution generates a local feature F_{loc}. The DSA output is

$$\mathrm{DSA}(X) = F_{loc} \odot A_{\mathrm{DSA}} + F_{loc} \tag{A.8}$$

This corresponds directly to the directional pooling, sigmoid attention, and residual feature enhancement implemented in DSA.

A.3 SFSSB

SFSSB is used at 16^3 and 8^3, where the reduced spatial resolution makes global state-space processing computationally tractable.

For input X, the conditioned feature U is

$$U = \Phi_b\left(\mathrm{GN}(X); n\right) \tag{A.9}$$

SMDG and FACA process the same conditioned representation in parallel:

$$Z = X + \mathrm{SMDG}\left(U\right) + \mathrm{FACA}\left(U\right) \tag{A.10}$$

A second conditioned normalization followed by a convolutional MLP produces

$$X_{out} = Z + \mathrm{MLP}\left(\Phi_b\left(\mathrm{GN}(Z); n\right)\right) \tag{A.11}$$

The MLP uses an expansion ratio of four:

$$C \to 4C \to C$$

implemented using two $1 \times 1 \times 1$ convolutions with an intermediate GELU activation.

Within an individual SFSSB block, the same GroupNorm module and block-specific PBAM projection are reused at the two conditioning locations.

A.4 SMDG

For

$$U \in \mathbb{R}^{B\times C\times D\times H\times W}$$

SMDG first generates three one-dimensional directional descriptors:

$$u_D = \mathrm{Avg}_{H,W}(U), u_H = \mathrm{Avg}_{D,W}(U), u_W = \mathrm{Avg}_{D,H}(U) \tag{A.12}$$

Each descriptor contains C channels and is divided into four equal channel groups. Here, a indexes the retained axis, *j* indexes the channel group, and $u_a{}^j$ denotes group j of descriptor u_a:

$$u_a = \left[u_a^{(1)}, u_a^{(2)}, u_a^{(3)}, u_a^{(4)}\right],\ a \in \{D, H, W\}$$

Each group *j* is processed using a depthwise 1D convolution with kernel size k_j:

$$k_j \in \{3, 5, 7, 9\}$$

The same four depthwise convolutional operators are shared across the depth, height, and width branches. Separate GroupNorm layers are retained for each spatial axis. GN_a denotes the axis-specific normalization layer, DWConv a depthwise convolution, and Concat channel concatenation. The directional gates A_a ($a \in \{D, H, W\}$) follow Section 3.2 and are given by

$$A_a = \sigma\left(\mathrm{GN}_a\left(\mathrm{Concat}_j\left[\mathrm{DWConv}_{k_j}\left(u_a^{(j)}\right)\right]\right)\right),\quad a \in \{D, H, W\} \tag{A.13}$$

After broadcasting A_D, A_H, and A_W to the original volumetric dimensions, the SMDG output is

$$\mathrm{SMDG}(U) = U \odot A_D \odot A_H \odot A_W \tag{A.14}$$

Thus, the module combines multi-scale directional filtering with parameter sharing across the three spatial axes. Here $\odot$ denotes element-wise multiplication with broadcasting.

A.5 FACA

FACA first applies Group Normalization and a $1 \times 1 \times 1$ projection:

$$V = \mathrm{Conv}_{1\times1\times1}\left[\mathrm{GN}(U)\right] \tag{A.15}$$

The projected feature is converted to float32 and transformed to the Fourier domain using an orthonormally normalized 3D FFT, producing the spectrum F_{freq}:

$$F_{freq} = \mathcal{F}_{3D}(V) \tag{A.16}$$

The real and imaginary components are concatenated along the channel dimension to produce F_{freq_c}:

$$F_{freq_c} = \mathrm{Concat}[\mathrm{Re}(F_{freq}),\ \mathrm{Im}(F_{freq})]$$

The three-dimensional frequency lattice is then flattened into a sequence:

$$Q \in \mathbb{R}^{B\times(DHW)\times 2C} \tag{A.17}$$

The Mamba-3 configuration (Lahoti et al., 2026) uses model dimension d_{model}, state dimension d_{state}, and expansion factor expand:

$$d_{model} = 2C,\ d_{state} = 16,\ \mathrm{expand} = 2$$

The head dimension d_{head} is

$$d_{head} = \max(16, \min(64, \frac{d_{model}}{2})) \tag{A.18}$$

For the SFSSB channel dimensions used in this network ($C = 64$ and $C = 128$), this gives

$$d_{head} = 64$$

Mamba-3 is used in SISO mode. The MIMO rank is therefore not applicable. The kernel chunk size follows the official default (chunk_size = 64). The Mamba-3 output is reshaped back to the 3D frequency lattice and separated into real and imaginary components to construct a complex representation $\overline{F}_{freq}$. The channel-wise gain from Section 3.2 has shape

$$w_f \in \mathbb{R}^{1\times C\times 1\times 1\times 1}$$

The gain-modulated spectrum $F_{freq,w}$ is

$$F_{freq,w} = \overline{F}_{freq} \odot w_f \tag{A.19}$$

Because w_f is broadcast over the three frequency dimensions, it performs channel-wise modulation rather than frequency-bin-specific weighting.

The real component of the inverse 3D Fourier transform is retained and locally refined:

$$S = \mathrm{Conv}_{3\times 3\times 3}\left[\mathrm{Re}\left(\mathcal{F}_{3D}^{-1}\left(F_{freq,w}\right)\right)\right] \tag{A.20}$$

The channels of S are divided into four equal groups, denoted by S_j:

$$S = [S_1, S_2, S_3, S_4]$$

Their element-wise mean $\overline{S}$ is

$$\overline{S} = \frac{1}{4}\sum_{j=1}^{4} S_j \tag{A.21}$$

The modulation feature is then generated as

$$G = \mathrm{Conv}_{1\times 1\times 1}\left[\mathrm{GELU}\left(\overline{S}\right)\right] \tag{A.22}$$

Finally,

$$\mathrm{FACA}(U) = S \odot G \tag{A.23}$$

This formulation exactly reflects the implemented FFT-Mamba-3-IFFT pathway and subsequent channel-group modulation.

The real projection is part of the operator definition. Let $H(k)$ denote the processed complex spectrum, with frequency indices interpreted modulo the grid dimensions. Define its Hermitian projection as $\mathcal{P}_H H(k) = [H(k) + \overline{H(-k)}]/2$. Then

$$\mathrm{Re}\big[\mathcal{F}^{-1}(H)\big] = \mathcal{F}^{-1}[\mathcal{P}_H H]\,.$$

Thus, the real-valued output does not require the learned spectrum to preserve conjugate symmetry. The discarded imaginary component is not assumed to be numerical roundoff or to contain no useful information. Appendix E.9 quantifies both imaginary-to-real norm ratios and Hermitian residuals without changing the reconstruction pathway. For its orthonormal FFT, the diagnostics satisfy $r_H = 2\rho_{\mathrm{imag}}/\sqrt{1+\rho_{\mathrm{imag}}^2}$ when ϵ is negligible, and therefore are not independent evidence.

A.6 PHOTON-BUDGET CONDITIONING

The reconstruction model is trained jointly over eight photon budgets:

$$10^2,\ 10^3,\ 10^4,\ 5\times10^4,\ 10^5,\ 2\times10^5,\ 5\times10^5,\ 10^6$$

The index-to-budget mapping follows Section 3.3.

The embedding lookup is

$$c_n = E[n],\ c_n \in \mathbb{R}^{64} \tag{A.24}$$

The block-specific projection is

$$[\gamma_b, \beta_b] = P_b(c_n) = W_b c_n + b_b \tag{A.25}$$

Where

$$\gamma_b, \beta_b \in \mathbb{R}^{C_b}$$

After broadcasting over the three spatial dimensions, the normalized feature F is modulated according to

$$\Phi_b(F; n) = F \odot (1 + \gamma_b) + \beta_b \tag{A.26}$$

Both the weights and biases of every PBAM projection are initialized to zero. Therefore,

$$\gamma_b = 0,\ \beta_b = 0$$

at initialization, and the modulation initially reduces to the identity mapping.

Each ConvResBlock and SFSSB module has an independent PBAM projection. Within a given SFSSB block, the same projection is reused at its two conditioning locations.

B DETAILED TRAINING OBJECTIVE

B.1 VARIANCE-STABILIZING TRANSFORMATION

For both finite-photon inputs and MC references, preprocessing first computes $u = \ln(1+e^8 a)$ from the stored MC absorption values a, then resamples the log volume to 64^3 using trilinear interpolation (`align_corners=False`). The resulting volumes are denoted by x and y, respectively. Training applies the Anscombe transformation (Anscombe, 1948) to these log-preprocessed volumes

$$\widetilde{x} = \mathcal{A}(x) = 2\sqrt{\max(x, 0) + \frac{3}{8}} \tag{B.1}$$

and

$$\widetilde{y} = \mathcal{A}(y)$$

At inference, the reconstructed transformed-domain output is mapped back to the log-preprocessed volume domain using the algebraic inverse

$$\mathcal{A}^{-1}(z) = \max\left(\left(\frac{z}{2}\right)^2 - \frac{3}{8}, 0\right) \tag{B.2}$$

No additional intensity normalization is applied by the current training pipeline after the Anscombe transformation.

B.2 Weighted Transformed-Domain MSE

Following Section 3.1, let $B = |\mathcal{B}|$ denote the mini-batch size. The per-sample transformed-domain error, photon-budget weight, and batch loss are

$$\mathrm{MSE}_i = \mathrm{mean}_{C,D,H,W}\left(\left(\widehat{\widetilde{y}}_i - \widetilde{y}_i\right)^2\right) \tag{B.3}$$

$$w(n_i) = \begin{cases} 1.10, & n_i \geq 3, \\ 1.00, & n_i < 3. \end{cases} \tag{B.4}$$

$$\mathcal{L}_{wMSE} = \frac{1}{B}\sum_{i=1}^{B} w\left(n_i\right)\mathrm{MSE}_i \tag{B.5}$$

B.3 Three-Dimensional Structural Loss

Structural consistency is encouraged using a three-dimensional extension of the standard SSIM formulation (Wang et al., 2004), calculated directly in the transformed domain.

The implementation uses a separable Gaussian window of size

$$11 \times 11 \times 11$$

with Gaussian-window standard deviation σ_G:

$$\sigma_G = 1.5$$

For local means μ_1, μ_2, variances σ_1^2, σ_2^2, and covariance σ_{12}, subscripts 1 and 2 refer to the transformed prediction and reference, respectively. C_1 and C_2 are stability constants. The implemented SSIM is

$$\mathrm{SSIM} = \frac{\left(2\mu_1\mu_2 + C_1\right)\left(2\sigma_{12} + C_2\right)}{\left(\mu_1^2 + \mu_2^2 + C_1\right)\left(\sigma_1^2 + \sigma_2^2 + C_2\right)} \tag{B.6}$$

where

$$C_1 = 0.01^2,\ C_2 = 0.03^2 \tag{B.7}$$

The structural loss averages the SSIM map over the mini-batch:

$$\mathcal{L}_{\mathrm{SSIM}} = 1 - \mathrm{mean}\left[\mathrm{SSIM}\left(\widehat{\widetilde{y}}, \widetilde{y}\right)\right] \tag{B.8}$$

Its coefficient depends on the proportion of low-photon samples in the current mini-batch. The indicator $\mathbf{1}_{\{\cdot\}}$ equals one when its condition holds and zero otherwise:

$$\lambda_s\left(\mathcal{B}\right) = 0.2 + 0.2\left[\frac{1}{B}\sum_{i=1}^{B}\mathbf{1}_{\{n_i \leq 2\}}\right] \tag{B.9}$$

Hence,

$$0.2 \leq \lambda_s\left(\mathcal{B}\right) \leq 0.4$$

This is specifically a mini-batch-level adaptive weight. The implementation computes one SSIM value for the mini-batch and scales it according to the proportion of samples belonging to the 10^2, 10^3, and 10^4 photon budgets.

B.4 High-Photon Log-Domain Fidelity

For samples with photon budgets of at least 2×10^5, let $\mathcal{H}$ be the set of eligible sample indices. Here, $|\cdot|$ denotes set cardinality, C = 1 for the single-channel absorption fields, and $\|\cdot\|_2$ is the Euclidean norm over all voxels and channels:

$$\mathcal{H} = \{i : n_i \geq 5\} \tag{B.10}$$

The network prediction and 10^8-photon MC reference are returned to the log-preprocessed volume domain using Equation (B.2). The subscript raw in $\mathcal{L}_{raw}$ denotes the stored volume before Anscombe, not physical absorption values. This log-domain reconstruction term is

$$\mathcal{L}_{raw} = \frac{1}{|\mathcal{H}|} \sum_{i \in \mathcal{H}} \frac{1}{CDHW} \| \mathcal{A}^{-1}\left(\widehat{\widetilde{y}}_i\right) - \mathcal{A}^{-1}\left(\widetilde{y}_i\right) \|_2^2 \tag{B.11}$$

This term is included only when at least one eligible sample occurs in the current mini-batch.

B.5 Highest-Photon Identity Regularization

An additional identity constraint is applied only to the 10^6-photon budget. The eligible sample-index set is $\mathcal{I}$:

$$\mathcal{I} = \{i : n_i = 7\} \tag{B.12}$$

The identity term is evaluated in the transformed domain:

$$\mathcal{L}_{id} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \frac{1}{CDHW} \| \widehat{\widetilde{y}}_i - \widetilde{x}_i \|_2^2 \tag{B.13}$$

It discourages excessive modification when the finite-photon estimate is already comparatively stable. The term is omitted if the current mini-batch contains no 10^6-photon samples.

Eqs. (B.11) and (B.13) are evaluated only over their corresponding eligible sample subsets.

B.6 Complete Training Objective

Combining the preceding components, the implemented objective is

$$\mathcal{L} = \mathcal{L}_{wMSE} + \lambda_s(\mathcal{B})\, \mathcal{L}_{\text{SSIM}} + 0.08\, \mathcal{L}_{raw} + 0.03\, \mathcal{L}_{id} \tag{B.14}$$

Thus, the objective combines photon-weighted transformed-domain reconstruction fidelity, local volumetric structural consistency, high-photon log-domain fidelity, and conservative reconstruction at the highest photon budget.

C Training and Evaluation Protocol

C.1 Dataset Partitioning

The data comprise 405 randomly selected underlying tissue models, each simulated at eight photon-budget levels. To prevent different photon-budget realizations of the same underlying tissue model from appearing in different subsets, the train/validation/test partition is generated over the 405 underlying tissue models and then propagated identically across all eight photon budgets.

The 405 underlying tissue models are partitioned as

$$405 = 243 \text{ training} + 81 \text{ validation} + 81 \text{ testing}. \tag{C.1}$$

The same model-level assignment, corresponding to a 243/81/81 training/validation/test split, is subsequently propagated identically across all eight photon budgets. Thus, all eight finite-photon realizations of a given underlying tissue model are confined to the same subset.

A fixed random seed of 42 is used for data partitioning.

C.2 Optimization Settings

The model is optimized using AdamW with a base learning rate

$$\eta_0 = 3 \times 10^{-4}$$

and weight decay

$$10^{-4}$$

Training is performed for up to 200 epochs with a batch size of 12.

The first five epochs use linear warm-up, followed by cosine learning-rate decay. The multiplicative learning-rate schedule implemented in the training code is

$$g_{sched}(e) = \begin{cases} \frac{e+1}{5}, & e < 5, \\ \frac{1}{2}\left[1 + \cos\left(\pi \min\left[1, \frac{e-5}{195}\right]\right)\right], & e \geq 5, \end{cases} \tag{C.2}$$

Here, e is the zero-based scheduler epoch index, g_{sched} is the multiplicative schedule, and η_e is the resulting learning rate:

$$\eta_e = \eta_0 \, g_{sched}(e) \tag{C.3}$$

With ∇_θ denoting the gradient with respect to θ, the gradient norms are clipped to

$$\| \nabla_\theta \mathcal{L} \|_2 \leq 1.0$$

before each optimizer update.

The detailed optimization settings are summarized in Table C1.

Table C1: Training configuration.

Setting	**Value**
Optimizer	AdamW
Base learning rate	3×10^{-4}
Weight decay	1×10^{-4}
Maximum epochs	200
Batch size	12
Warm-up	5 epochs
LR schedule after warm-up	Cosine decay
Gradient clipping	1
Mixed-precision training	Disabled
Model-selection criterion	Minimum validation loss
Data-split random seed	42

Automatic mixed precision is disabled in the reported configuration.

C.3 Model Selection and Inference

Validation loss is evaluated after each epoch. Model parameters are saved whenever the validation loss reaches a new minimum, and the checkpoint corresponding to the lowest validation loss is used for final evaluation.

During inference, the stored log-preprocessed finite-photon volume is transformed as

$$\widetilde{x} = \mathcal{A}(x). \tag{C.4}$$

The trained network predicts

$$\widehat{\widetilde{y}} = \widetilde{x} + \mathcal{R}_\theta(\widetilde{x}, n) \tag{C.5}$$

after which the reconstruction is returned to the log-preprocessed volume domain:

$$\widehat{y} = \mathcal{A}^{-1}\left(\widehat{\widetilde{y}}\right) \tag{C.6}$$

Inverse Anscombe returns the stored log-preprocessed representation. The logarithmic preprocessing is not inverted before evaluation.

C.4 Reconstruction Metrics

Unless explicitly labeled otherwise, SSIM, PSNR, and RMSE use the matched 10^8-photon MC reference in the log-preprocessed volume domain after inverse Anscombe. The additional PSNR evaluation in Tables 5 and E8 instead uses an independently simulated 10^9-photon reference. Both PSNR evaluations retain the stored log scale. Raw MC denotes the unreconstructed input on that scale, as do the reconstruction maps in the main text. The separately labeled inverse-transformed field evaluation in Table E11 uses the definitions below.

The reference-specific dynamic range L_i is

$$L_i = \max(y_i) - \min(y_i) \tag{C.7}$$

If the reference volume is constant, the implementation sets

$$L_i = 1$$

PSNR is calculated as follows, with voxel-wise MSE:

$$\mathrm{PSNR}_i = 10 \log_{10}\left(\frac{L_i^2}{\mathrm{MSE}(\widehat{y}_i, y_i)}\right) \tag{C.8}$$

RMSE is

$$\mathrm{RMSE}_i = \sqrt{\frac{1}{DHW} \parallel \widehat{y}_i - y_i \parallel_2^2} \tag{C.9}$$

SSIM is also calculated using the reference-specific dynamic range L_i. Metrics are first calculated separately for each complete 64^3 volume and then averaged over the test configurations within each budget (81 for the main split; Appendix F gives the separate physical-regime splits). The same PSNR implementation is used for both reference budgets, with L_i computed from the reference used in that evaluation. Consequently, PSNR against 10^9 need not be smaller than PSNR against 10^8: both the reference error and its dynamic range can change.

The same procedure is applied to both the log-preprocessed finite-photon inputs and the reconstructed outputs, enabling direct quantification of the improvement produced by the reconstruction network.

Inverse-transformed field metrics. Stored inputs and references are saved before Anscombe transformation; predictions are saved after inverse Anscombe. All therefore share $u = \ln(1 + e^8 a)$. Applying $A = \mathrm{expm1}(u)/e^8$ gives the inverse-transformed absorption field on the common

64^3 evaluation grid. For a reconstructed field $\widehat{A}_i$, reference A_i, and depth profile $p_i(z) = \sum_{x,y} A_i(x, y, z)$, we report

$$E_{\mathrm{int},i} = 100\frac{\left|\sum_{x,y,z} \widehat{A}_i - \sum_{x,y,z} A_i\right|}{\sum_{x,y,z} A_i}, \qquad E_{\mathrm{depth},i} = 100\frac{\|\widehat{p}_i - p_i\|_2}{\|p_i\|_2}.$$

Raw MC is evaluated by the same formulas. Equal voxel-volume and plane-area factors cancel in these ratios. Errors are averaged over 81 configurations within each budget and equally over the eight budgets for the macro-average. The depth axis is the first spatial array dimension. Because preprocessing applies log transformation before resampling, this inverse-transformed field is not generally identical to directly resampling the native-grid absorption field. These reference-relative errors measure neither incident-energy balance nor complete local field accuracy.

D Experimental Setup Details

D.1 Benchmark Design and Motivation

We construct a controlled three-dimensional photon-transport benchmark based on multilayer pigmented skin with an embedded BCC-like lesion. The purpose of this model is not to simulate a complete clinical diagnostic setting, but to provide a structured and physically motivated reconstruction problem in which finite-photon sampling errors interact with heterogeneous tissue geometry.

This setting is useful for evaluating finite-photon MC reconstruction for three reasons. First, the melanin-rich superficial layer introduces strong absorption near the tissue surface, producing pronounced depth-dependent redistribution of photon energy. Second, the layered skin structure generates extended spatial dependencies across tissue interfaces. Third, the irregular embedded lesion introduces localized absorption and scattering perturbations whose boundaries are spatially coupled to the surrounding tissue. The resulting absorption fields therefore contain both local high-gradient structures and long-range volumetric transport patterns, providing a more stringent reconstruction benchmark than spatially homogeneous phantoms.

The benchmark is used solely to evaluate whether a learned model can reconstruct a matched 10^8-photon MC reference from a finite-photon MC absorption field under identical geometry and optical properties. The reconstruction network is not intended to replace MC as the physical forward solver or to provide a diagnostic prediction of BCC.

D.2 Monte Carlo Tissue Model

Photon transport was simulated using MCmatlab (Marti et al., 2018) in a cubic tissue domain of

$$0.64 \times 0.64 \times 0.64\ cm^3$$

discretized into

$$64 \times 64 \times 64$$

voxels. The isotropic voxel spacing was therefore

$$0.01\ cm = 0.1\ mm$$

All simulations were performed at a wavelength of $633\ nm$.

The background tissue consisted of five planar regions: epidermis, a melanin-rich layer, dermis, subcutaneous fat, and muscle. An irregular BCC-like lesion was embedded from the superficial skin into the dermal region rather than modeled as an additional planar layer. Lesion morphology was generated by combining three-dimensional Gaussian components with varying spatial positions, sizes, and weights, producing non-planar lesion boundaries and spatially varying lesion influence.

The optical and geometric parameters used to construct the candidate models are summarized in Table D1. They are literature-grounded benchmark design values rather than subject-specific measurements or a one-to-one transcription from a single specimen.

Normal-skin absorption and reduced-scattering ranges were informed by layered-skin modeling and measurements of epidermis, dermis, subcutaneous fat, and muscle (Meglinski & Matcher, 2002; Bashkatov et al., 2005; 2011; Shimojo et al., 2020). Refractive-index values were selected within reported tissue ranges near 633 nm (Bolin et al., 1989; Bashkatov et al., 2011). Pigmentation-related variation motivated the melanin absorption sweep (Setchfield et al., 2024), while lesion optical-property and invasion-depth ranges were informed by measurements of normal and cancerous skin, in vivo nonmelanoma skin-cancer optical properties near 630 nm, and a large histopathologic BCC series (Salomatina et al., 2006; Saager et al., 2013; Pyne et al., 2018).

The exact discrete sweep levels, Gaussian morphology scales, and sampled invasion positions are controlled benchmark design choices.

Table D1: Optical and geometric parameters of the multilayer skin-BCC MC benchmark at 633 nm.

Component	**Thickness**	μ_a **(cm^{-1})**	μ_s' **(cm^{-1})**	**g**	n_{ref}	**Dataset treatment**
Epidermis	0.10 mm	0.5	42.5	0.875	1.425	Fixed
Melanin layer	0.02 mm	8, 14, 20	37.5	0.85	1.41	μ_a swept, 3 levels
Dermis	1.50 mm	0.2, 0.4, 0.6	27.5	0.875	1.385	μ_a swept, 3 levels
Subcutaneous fat	3.00 mm	0.06	12.5	0.85	1.45	Fixed
Muscle	1.78 mm	1.0	7.5	0.935	1.375	Fixed
BCC-like lesion	Irregular	0.5, 1.0, 1.5	10, 14, 18	0.925	1.39	μ_a and μ_s' swept
Lesion scale (FWHM)	Gaussian FWHM = 0.5, 0.8, 1.2, 1.6 mm	—	—	—	—	Gaussian FWHM = 0.5, 0.8, 1.2, 1.6 mm
Lesion invasion depth	10 sampled positions	—	—	—	—	10 sampled positions

Here, μ_a denotes the absorption coefficient, μ'_s the reduced scattering coefficient, g the anisotropy factor, and n_{ref} the refractive index. The tissue subscripts mel, dermis, and lesion identify melanin, dermis, and the BCC-like lesion; s_{lesion} denotes the lesion morphology scale expressed as full width at half maximum (FWHM), and d_{inv} the lesion invasion depth. The parameter sweep varied six factors:

$$\mu_{a,mel},\ \mu_{a,dermis},\ \mu_{a,lesion},\ \mu'_{s,lesion},\ s_{lesion},\ d_{inv}$$

with respectively

$$3,\ 3,\ 3,\ 3,\ 4,\ 10$$

settings. The full Cartesian parameter space therefore contained

$$3 \times 3 \times 3 \times 3 \times 4 \times 10 = 3240$$

candidate tissue configurations.

D.3 Illumination and Boundary Conditions

We injected a normally incident pencil beam at the center of the top face of a cubic domain of side length $0.64\ cm$. The source was $0.001\ cm$ outside that face, with polar and azimuthal angles of zero. The beam had no finite diameter, Gaussian waist, numerical aperture, or top-hat profile.

We retained refractive-index mismatch between tissue regions, which produced Fresnel reflection at internal interfaces. Photons that reached the outer faces of the domain escaped. MCmatlab computed a normalized fluence-rate volume; the corresponding normalized three-dimensional absorption field was log-preprocessed and resampled as specified in Appendix B.1 to form the stored reconstruction target.

D.4 Candidate Subsampling and Photon-Budget Pairing

Running the complete photon-budget sweep over all 3,240 candidate tissue configurations would substantially increase simulation storage and training cost. We therefore selected a fixed subset of 405 tissue configurations before constructing the reconstruction dataset.

The selected configuration IDs were sampled from the full pool of 3,240 candidate tissue configurations using MATLAB randperm with a fixed Twister seed of 23. Importantly, sampling was performed at the underlying 3D model level, and the same 405 configuration IDs were subsequently used at every photon budget.

For each selected tissue model, finite-photon absorption volumes were generated using $10^2,\ 10^3,\ 10^4,\ 5\times10^4,\ 10^5,\ 2\times10^5,\ 5\times10^5,\ 10^6$ photons. A corresponding 10^8-photon simulation served as the MC reference. For an underlying model m, the reconstruction pairs are therefore

$$x_m^{(n)}, y_m$$

where $x_m^{(n)}$ denotes the log-preprocessed finite-photon volume at photon-budget index n, while y_m denotes the log-preprocessed 10^8-photon MC reference generated using exactly the same geometry and optical-property configuration.

Each underlying model contributes eight reconstruction pairs:

$$405 \times 8 = 3240$$

reconstruction pairs in total. These 3,240 reconstruction pairs should not be interpreted as 3,240 independent candidate tissue configurations; they correspond to 405 selected tissue configurations observed at eight different finite-photon budgets.

D.5 Leakage-Free Dataset Partitioning

Because the same tissue configuration is simulated at multiple photon budgets, randomly splitting individual finite-photon volumes would introduce information leakage: different photon-budget realizations of the same underlying geometry could otherwise occur in both training and test sets.

To prevent this, splitting is performed first over the 405 underlying model indices. Using a fixed random seed of 42, the models are divided into 243 training models, 81 validation models, and 81 test models, corresponding to a $60\%/20\%/20\%$ partition.

The resulting underlying-model assignment is then propagated identically to all eight photon-budget levels. Therefore, if model m is assigned to the test set, all eight finite-photon realizations associated with model m are restricted to the test set.

The effective numbers of reconstruction pairs are consequently $243\times8 = 1944$ for training, $81\times8 = 648$ for validation, and $81 \times 8 = 648$ for testing.

The proposed volumetric model operates directly on complete three-dimensional volumes without slice decomposition. For the controlled slice-wise comparison described in Appendix E.2, XZ slices

are extracted only after the underlying-model partitioning defined above, ensuring that slices from the same 3D model never cross the training, validation, and test subsets.

D.6 Unified Multi-Budget Training Protocol

All eight photon-budget levels are pooled during training of the unified reconstruction model. Each sample consists of a finite-photon volume together with its discrete photon-budget index

$$(x_i, n_i)$$

and the corresponding 10^8-photon MC reference y_i.

The photon-budget index is mapped as

$$0 \to 10^2$$

$$1 \to 10^3$$

$$2 \to 10^4$$

$$3 \to 5 \times 10^4$$

$$4 \to 10^5$$

$$5 \to 2 \times 10^5$$

$$6 \to 5 \times 10^5$$

$$7 \to 10^6$$

Visualization masks for Figures 4 and 5 are reconstructed on an endpoint-inclusive grid with 64 samples per axis over $[-0.32, 0.32]$ cm (spacing $0.64/63$ cm), distinct from the MC solver voxel-center grid. These stored masks determine display-plane selection and lesion overlays. No lesion mask, tissue-label map, or optical-property vector is provided to the reconstruction network. The model receives only the finite-photon 3D volume and its photon-budget index. Thus, improvements in lesion-associated structure arise from reconstruction of the absorption field rather than explicit access to lesion geometry.

Both log-preprocessed finite-photon inputs and 10^8-photon MC references are transformed using the Anscombe-style transformation defined in Section 3.4 before being passed to the network. The same underlying split is used for all photon-budget levels.

Detailed loss definitions and optimization settings are given in Appendices B and C, respectively.

D.7 Evaluation Protocol

Network predictions are first converted from the variance-stabilized domain back to the log-preprocessed volume domain using the inverse transformation defined in Section 3.4. Evaluation is then performed independently for every 3D test volume.

The reference-specific dynamic range, constant-reference fallback, PSNR and RMSE follow Appendix C.4, Eqs. (C.7)–(C.9). SSIM is computed using the same reference-specific dynamic range.

Metrics are computed first at the individual-volume level and then averaged separately within each photon-budget group. $M_i^{(n)}$ denotes a selected metric (PSNR, SSIM, or RMSE) for test volume i at index n; $\bar{M}^{(n)}$ is its average, and N_{test} is the number of test volumes:

$$\bar{M}^{(n)} = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} M_i^{(n)}, \; N_{test} = 81 \tag{D.1}$$

The same procedure is applied to both the log-preprocessed finite-photon inputs and the reconstructed outputs, enabling direct measurement of reconstruction improvement at each photon budget.

D.8 Controlled Comparison Protocol

All architectural reconstruction baselines use the same 405 underlying tissue models and the identical 243/81/81 model-level split. The photon-specific comparison uses a separate training-set construction to match the number of training pairs per model while retaining an identical held-out evaluation set.

Specifically, the 81 validation and 81 test configuration IDs from the unified-model split are fixed for all three reconstruction strategies. For each photon-specific model, these 162 held-out configuration IDs are first excluded from the full pool of 3,240 candidate tissue configurations, leaving 3,078 eligible configurations. A total of 1944 candidate tissue configurations are then selected for training at that photon budget. Each photon-specific model therefore uses 1944 single-budget training pairs, matching the 1944 training pairs used to train the unified model, which contains 243 selected underlying tissue models observed at eight photon budgets (243×8=1944).

This design gives each photon-specific model substantially greater unique training-geometry diversity than the unified model while ensuring that all strategies are validated and evaluated on exactly the same held-out configurations. No validation or test geometry is included in the training set of any photon-specific model.

D.9 Baseline Implementations

All reconstruction baselines use the same paired log-preprocessed finite-photon and 10^8-photon MC absorption fields, the same 405 selected tissue configurations, and the identical 243/81/81 model-level training/validation/test split defined in Appendix D.5. Each baseline is trained jointly across all eight photon budgets using a single model rather than separate photon-specific models. For slice-wise methods, the model-level split is completed before slice extraction, and the reconstructed slices are reassembled into complete 64^3 volumes before evaluation. Consequently, PSNR, SSIM, and RMSE are evaluated using the same volume-level protocol for all methods.

The baseline implementations and task adaptations are summarized in Table D2 and Table D3. Source architectures were retained wherever compatible with the present reconstruction task, while operators specific to their original applications were removed or adapted when necessary. The common comparison protocol controls the reconstruction data, underlying-model partitioning, multi-budget training setting, and volume-level evaluation; method-specific transformations, losses, and optimization settings are retained where required by the corresponding implementation. Parameter counts denote the trainable parameters of the task-adapted models actually used in the reported experiments. Mamba-based baselines are identified by the cited source papers together with the Mamba kernel versions used at training (VEMamba: mamba_ssm 2.2.5 with causal_conv1d 1.5.2; DenoMamba: mamba-ssm 1.1.3).

ReconFormer and DenoMamba are evaluated as slice-wise baselines because their source architectures are two-dimensional; the remaining baselines operate volumetrically. For both slice-wise methods, the underlying-model split is performed before slice extraction, preventing slices from the same 3D volume from crossing the training, validation, and test subsets. At inference, all 64 reconstructed slices are reassembled before the common 3D metric calculation.

The use of photon-budget information also differs across methods. The proposed model uses block-specific Photon-Budget Affine Modulation (PBAM) to condition intermediate representations ex-

plicitly on photon budget. VEMamba receives the photon-budget index through its adapted gated-residual pathway, whereas the remaining baselines do not receive photon budget as input to their forward reconstruction networks; where applicable, the index is used only for loss weighting. Thus, the comparison controls the data, model-level partitioning, multi-budget training setting, and volume-level evaluation without forcing different architectures to share the proposed model's conditioning mechanism or training objective.

MC-Fourier follows the published Fourier-reconstruction design (Fu et al., 2025) and is adapted to the same 64^3 finite-photon MC reconstruction task. The reported 3.28 M parameter count corresponds to the task-adapted implementation used in the present experiments.

The 3D U-Net and DnCNN-U-Net baselines follow the U-Net, 3D U-Net, DnCNN, and MC-denoising formulations (Ronneberger et al., 2015; Çiçek et al., 2016; Zhang et al., 2017; Raayai Ardakani et al., 2022). The remaining baselines follow 3D FNO/Neural Operator (Kovachki et al., 2023), MC-Fourier with an RCAN backbone (Fu et al., 2025; Zhang et al., 2018), ReconFormer (Guo et al., 2024), VEMamba (Gao & Gao, 2026), and DenoMamba (Öztürk et al., 2026).

Table D2: Source implementations and reconstruction settings of the evaluated baselines.

Method	**Source implementation**	**Reconstruction mode**	**Photon-budget use**	**Params**
3D U-Net (Ronneberger et al., 2015; Çiçek et al., 2016; Zhang et al., 2017; Raayai Ardakani et al., 2022)	MCX-Denoising-Pytorch; 3D residual U-Net	Volumetric 64^3	Unified eight-budget training; no photon-budget input to the network	5.97 M
DnCNN-U-Net (Ronneberger et al., 2015; Çiçek et al., 2016; Zhang et al., 2017; Raayai Ardakani et al., 2022)	MCX-Denoising-Pytorch; 17-layer 3D DnCNN followed by 3D U-Net	Volumetric 64^3	Unified eight-budget training; no photon-budget input to the network	7.64 M
3D FNO (Kovachki et al., 2023)	Official neuraloperator FNO implementation	Volumetric 64^3	Unified eight-budget training; no photon-budget input	18.89 M
MC-Fourier (Fu et al., 2025)	Published MC-Fourier design; RCAN reconstruction backbone	Volumetric 64^3	Unified eight-budget training; photon-budget index used only for loss weighting	3.28 M
ReconFormer (Guo et al., 2024)	Official ReconFormer implementation	Slice-wise 64×64; reassembled to 64^3	Unified eight-budget training; photon-budget index used only in the loss	1.14 M
VEMamba (Gao & Gao, 2026)	Official VEMamba implementation	Volumetric I/O 64^3; SSM backbone on a 32^3 grid	Unified eight-budget training; budget embedding controls an adapted gated residual	1.04 M
DenoMamba (Öztürk et al., 2026)	Official DenoMamba implementation	Slice-wise 64×64; reassembled to 64^3	Unified eight-budget training; no photon-budget input to the network	38.58 M

Table D3: Task adaptations and optimization protocols of the evaluated baselines.

Method	Task adaptation	Optimization protocol
3D U-Net	3D Conv/BN U-Net for MC photon denoising with residual noise prediction	AdamW, lr 3×10^{-4}, wd 1×10^{-4}; 5-epoch warm-up+cosine decay; max 200 epochs; Anscombe transform; MSE in the Anscombe domain; no additional intensity normalization; effective batch size 12; gradient clipping 1.0
DnCNN-U-Net	Original 3D DnCNN-U-Net cascade retained; both stages use residual reconstruction	AdamW, lr 3×10^{-4}, wd 1×10^{-4}; 5-epoch warm-up cosine decay; max 200 epochs; Anscombe transform; MSE objective; effective batch size 12
3D FNO	Four FNO layers with 16^3 Fourier modes and hidden width 32; same-resolution MC reconstruction	AdamW, lr 3×10^{-4}, wd 1×10^{-4}; StepLR with step = 60, $\gamma = 0.5$; 200 epochs; batch size 12; best validation-L_2 checkpoint
MC-Fourier	3D FFT $\rightarrow$ learnable Fourier masks $\rightarrow$ inverse FFT $\rightarrow$ RCAN; K = 11, C = 20, G = 5, B = 3; no photon-budget conditioning in the forward network	Anscombe transform; AdamW, lr 3×10^{-4}; 5-epoch warm-up+cosine decay; transformed-domain MSE+3D SSIM with level-aware weighting; max 200 epochs; batch size 12
ReconFormer	MRI-specific k-space and undersampling operations removed; applied to absorption-field slices with residual prediction. Split precedes slice extraction; 16 slices/volume are sampled per epoch and all 64 slices are reconstructed at inference	Anscombe transform; AdamW, lr 3×10^{-4}, wd 1×10^{-4}; 5-epoch warm-up+cosine decay; MSE+2D SSIM+level-aware weighting+log-domain MSE+identity regularization; batch size 12; early stopping on validation log-domain MSE
VEMamba	MoCo, VDIM, anisotropic upsampling, and PixelShuffle removed. Shallow Conv3d features remain at 64^3; a stride-2 space-to-depth layer maps them to a 32^3 Mamba-3D/RVMG backbone; depth-to-space restores 64^3, which is fused with the shallow skip and refined at native resolution. Same-resolution refers to $64^3\rightarrow64^3$ reconstruction rather than super-resolution. Photon budget enters through a gated residual rather than PBAM.	AdamW, lr 3×10^{-4}, wd 1×10^{-4}; 10-epoch warm-up + cosine decay; max 200 epochs; log-domain MSE+3D SSIM+level-aware MSE+identity regularization at 5×10^5 and 10^6; batch size 12
DenoMamba	Original hourglass architecture with spatial/channel state-space modules and global residual retained; split precedes slice extraction; 16 slices/volume sampled during training and all 64 reconstructed at inference	Anscombe transform; AdamW, lr 3×10^{-4}, wd 1×10^{-4}; 5-epoch warm-up+cosine decay; 2D MSE+SSIM; FP32; batch size 12; gradient clipping 0.5; early-stopping patience 20

E EXPERIMENTS

E.1 FULL COMPARISON

Table E1: Quantitative reconstruction performance across eight photon budgets: PSNR (dB).

Method	10^2	10^3	10^4	5×10^4	10^5	2×10^5	5×10^5	10^6
Raw MC	7.7095	8.9005	13.0527	19.5752	23.8414	28.6888	34.8453	38.6475
3D U-Net	29.5982	32.8294	34.3817	37.5082	39.6803	41.8611	44.0893	45.0474
DnCNN-U-Net	30.5611	33.8349	37.1473	41.5153	44.0514	46.4714	48.6938	49.3451
3D FNO	34.1840	37.2605	41.8793	46.8252	48.4199	49.2248	49.4023	49.4954
MC-Fourier	32.7156	37.7187	41.5689	44.4089	46.0165	47.2085	48.1499	48.4882
ReconFormer	29.2739	35.8789	42.5552	46.6833	48.6355	50.3348	51.9442	52.6980
VEMamba	33.9299	39.8138	42.6790	45.4025	46.8049	48.0736	49.1883	50.4737
DenoMamba	30.5064	35.8695	41.8741	45.7397	47.6949	49.3232	50.9989	51.7654
Proposed	34.0070	42.6318	47.7852	51.4834	53.1098	54.5029	56.2741	56.7556

Table E2: Quantitative reconstruction performance across eight photon budgets: SSIM.

Method	10^2	10^3	10^4	5×10^4	10^5	2×10^5	5×10^5	10^6
Raw MC	0.0055	0.0294	0.2096	0.5670	0.7163	0.8310	0.9235	0.9597
3D U-Net	0.9036	0.8912	0.8871	0.9525	0.9756	0.9868	0.9919	0.9934
DnCNN-U-Net	0.9233	0.8988	0.9355	0.9806	0.9916	0.9963	0.9980	0.9985
3D FNO	0.9875	0.9947	0.9981	0.9993	0.9996	0.9997	0.9997	0.9997
MC-Fourier	0.9820	0.9890	0.9933	0.9963	0.9972	0.9977	0.9980	0.9981
ReconFormer	0.8270	0.9450	0.9872	0.9949	0.9966	0.9974	0.9979	0.9981
VEMamba	0.9831	0.9885	0.9926	0.9959	0.9970	0.9978	0.9982	0.9987
DenoMamba	0.9237	0.9680	0.9873	0.9949	0.9971	0.9982	0.9988	0.9991
Proposed	0.9883	0.9963	0.9984	0.9992	0.9994	0.9996	0.9997	0.9997

Table E3: Quantitative reconstruction performance across eight photon budgets: RMSE.

Method	10^2	10^3	10^4	5×10^4	10^5	2×10^5	5×10^5	10^6
Raw MC	5.7601	5.0193	3.1121	1.4742	0.9053	0.5190	0.2551	0.1643
3D U-Net	0.4810	0.3218	0.2671	0.1863	0.1452	0.1130	0.0875	0.0784
DnCNN-U-Net	0.4295	0.2864	0.1943	0.1175	0.0878	0.0665	0.0515	0.0479
3D FNO	0.3064	0.2176	0.1273	0.0705	0.0587	0.0546	0.0552	0.0563
MC-Fourier	0.3499	0.1856	0.1174	0.0847	0.0706	0.0620	0.0560	0.0540
ReconFormer	0.5007	0.2277	0.1048	0.0650	0.0519	0.0427	0.0354	0.0325
VEMamba	0.3101	0.1451	0.1034	0.0754	0.0642	0.0556	0.0490	0.0424
DenoMamba	0.4390	0.2282	0.1134	0.0726	0.0580	0.0481	0.0398	0.0366
Proposed	0.3182	0.1078	0.0590	0.0381	0.0316	0.0268	0.0219	0.0206

E.2 SLICE-WISE RECONSTRUCTION

The slice-wise model was constructed as the dimensional 2D counterpart of the proposed volumetric model rather than as an independently designed architecture. It preserves the same encoder-decoder depth, channel configuration, skip connections, global residual formulation, SMDG, FACA, SFSSB, ConvResBlock, DSA, and photon-budget PBAM conditioning. The corresponding volumetric operators were replaced by their 2D counterparts, including 3D convolutions with 2D convolutions, three-axis spatial operations with two-axis operations, and 3D FFT/IFFT with 2D FFT/IFFT.

To maintain the same leakage-free protocol as the volumetric experiment, the 405 underlying 3D models were first partitioned into 243 training, 81 validation, and 81 test models using the same model-level split. XZ slices were extracted only after this partitioning, ensuring that all slices from

the same underlying model remained within the same subset. Each 64^3 volume contributed 64 two-dimensional slices. A single slice-wise model was trained jointly across all eight photon budgets using the same photon-budget embedding and block-specific PBAM conditioning as the volumetric model.

The slice-wise model used the same Anscombe variance-stabilizing transformation, photon-weighted transformed-domain reconstruction term, high-photon log-domain fidelity term, highest-photon identity regularization, and photon-dependent structural weighting. The structural term was implemented using 2D SSIM. Optimization used AdamW with a learning rate of 3×10^{-4}, weight decay of 1×10^{-4}, batch size 48, gradient clipping of 1, and a five-epoch linear warm-up followed by cosine decay. Training was performed for at most 200 epochs with an early-stopping patience of 20.

During inference, the independently reconstructed XZ slices corresponding to each test case were stacked back into a $64 \times 64 \times 64$ volume and then mapped to the log-preprocessed volume domain using the inverse Anscombe transformation. PSNR, SSIM, and RMSE were computed on the reconstructed 3D volumes using the same volume-level evaluation protocol as for the volumetric model. Table E4 reports the complete slice-wise reconstruction results across all eight photon budgets.

Table E4: Complete slice-wise reconstruction results across all eight photon budgets.

Photon budget	**PSNR↑**	**SSIM↑**	**RMSE↓**
10^2	30.2705	0.8882	0.4433
10^3	38.5075	0.9782	0.1685
10^4	45.2179	0.9954	0.0779
5×10^4	49.1572	0.9983	0.0494
10^5	50.2154	0.9989	0.0437
2×10^5	52.1064	0.9993	0.0351
5×10^5	54.2032	0.9995	0.0275
10^6	54.9655	0.9995	0.0251

E.3 Photon-Specific Reconstruction

To evaluate whether a single unified model can replace independently optimized models for different photon budgets, eight photon-specific models were trained separately from scratch, with one model assigned to each photon budget. Each photon-specific model uses the same spatial-frequency reconstruction architecture as the unified model without conditioning, but contains neither a photon-budget embedding nor PBAM modulation. Each model has 4,129,681 parameters (4.1297 M); the eight-model total is 33,037,448 (33.0374 M after rounding). The eight models use a common training configuration, and no photon-budget-specific hyperparameter search is performed.

As described in Appendix D.8, the validation and test configuration IDs are fixed to those used by the unified models. After excluding these held-out cases from the full pool of 3,240 candidate tissue configurations, each photon-specific model is trained using 1944 candidate tissue configurations from its corresponding photon budget. This matches the 1944 training pairs used to train the unified model while providing substantially greater geometry diversity to each photon-specific model.

Table E5 reports the complete photon-specific reconstruction results on the common test set. Because the input volumes and test geometries are identical to those used for unified reconstruction, the input PSNR, SSIM, and RMSE are identical across strategies.

Table E5: Photon-specific reconstruction performance on the common test set across eight photon budgets.

Photon budget	**PSNR↑**	**SSIM↑**	**RMSE↓**
10^2	34.3179	0.9894	0.2996
10^3	43.7236	0.9957	0.0937
10^4	47.3014	0.9978	0.0622
5×10^4	49.9416	0.9988	0.0455
10^5	51.9216	0.9991	0.0361
2×10^5	53.8927	0.9994	0.0288
5×10^5	56.1747	0.9997	0.0221
10^6	56.2670	0.9997	0.0218

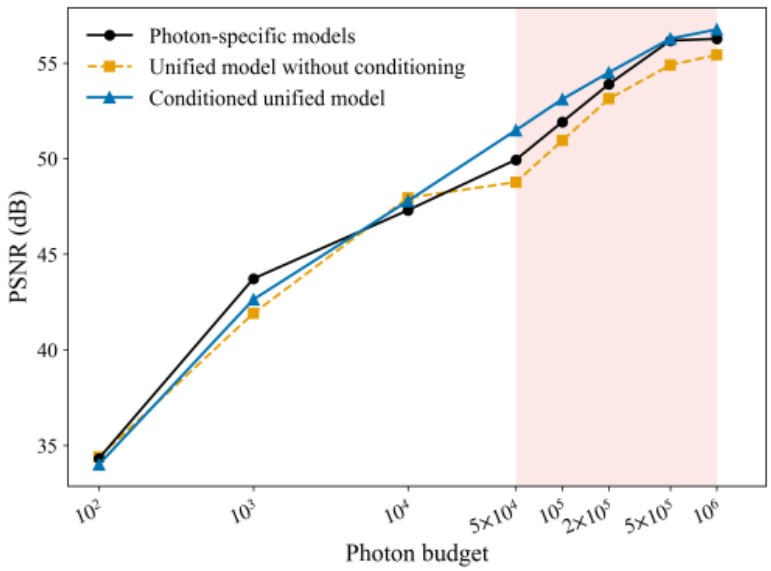

Figure E1: Per-budget PSNR for photon-specific, unified-unconditioned, and proposed reconstruction. Shading marks 5×10^4–10^6.

E.4 Complete Ablation Results Across Photon Budgets

All seven variants use the same 405 tissue configurations, 243/81/81 model-level split, unified eight-budget training, and evaluation protocol. All retain identical high-resolution DSA blocks at enc1, enc2, dec2, and dec1. Table 4 identifies the SMDG/FACA/PBAM settings; C and E use a capacity-matched convolutional control instead of FACA. A–F have no photon-budget embedding or PBAM, whereas G contains both. The exact parameter counts for A–G are 1,064,113; 1,070,769; 4,120,235; 4,123,025; 4,126,891; 4,129,681; and 4,209,233. Thus C/D and E/F each differ by only 2,790 parameters. Their respective reported FLOPs are 37.12, 37.12, 43.42, 43.40, 43.42, 43.40, and 43.40 G.

Capacity-matched convolutional control. At the FACA positions in enc3, the three bottleneck blocks, and dec3, the control uses two $3 \times 3 \times 3$ Conv3d layers (stride 1, padding 1, default bias) with an intermediate GELU. Its hidden width is 64 when the input has at most 64 channels, and 126 otherwise; enc3/dec3 have 64 input channels and the bottleneck has 128. The control contains no internal GroupNorm or residual connection. It retains the surrounding SFSSB normalization, external branch residual, and $1 \times 1 \times 1$ MLP. For the normalized branch input U,

$$\mathrm{ConvControl}(U) = \mathrm{Conv}^{(2)}_{3\times3\times3}\Big(\mathrm{GELU}\left(\mathrm{Conv}^{(1)}_{3\times3\times3}(U)\right)\Big).$$

Variant C has neither SMDG nor PBAM; E retains SMDG but no PBAM. The control is part of the ablation protocol, not the proposed network.

Counting and aggregation. Parameters count model tensor elements and were checked against the trained checkpoint. FLOPs use `torch.utils.flop_counter.FlopCounterMode` for one forward pass with input shape $(1, 1, 64, 64, 64)$, counting multiplication and addition separately. Unsupported FFT/IFFT and selective-scan operations are excluded, so these are partial operator counts, not total computational costs or evidence of equal execution cost. Per-budget metrics are averaged over 81 test volumes; Table 4 reports equal-weight macro-averages across the eight budgets. The per-budget entries below are rounded for display, so recomputing macro-averages from them can differ in the last decimal place.

Table E6: Complete component and capacity-controlled ablations across eight photon budgets. A–G follow Table 4; all variants retain the same high-resolution DSA.

Variant	PSNR↑	SSIM↑	RMSE↓	Variant	PSNR↑	SSIM↑	RMSE↓
Photon budget: 10^2				**Photon budget:** 10^5			
A	31.3334	0.9513	0.3986	A	49.1595	0.9988	0.0493
B	33.8219	0.9874	0.3327	B	49.9202	0.9990	0.0451
C	33.4388	0.9866	0.3311	C	49.4395	0.9988	0.0478
D	33.4879	0.9866	0.3356	D	50.7365	0.9991	0.0416
E	33.8300	0.9870	0.3326	E	50.0112	0.9987	0.0451
F	34.4066	0.9869	0.3100	F	50.9433	0.9992	0.0402
G	34.0070	0.9883	0.3182	G	53.1098	0.9994	0.0316
Photon budget: 10^3				**Photon budget:** 2×10^5			
A	40.8544	0.9939	0.1295	A	50.4841	0.9990	0.0425
B	42.3356	0.9950	0.1104	B	51.4316	0.9992	0.0381
C	39.2604	0.9928	0.1610	C	51.0050	0.9991	0.0402
D	41.2910	0.9944	0.1277	D	52.4471	0.9994	0.0344
E	40.4793	0.9936	0.1403	E	51.1917	0.9989	0.0398
F	41.9079	0.9951	0.1185	F	53.1281	0.9994	0.0315
G	42.6318	0.9963	0.1078	G	54.5029	0.9996	0.0268
Photon budget: 10^4				**Photon budget:** 5×10^5			
A	45.4642	0.9972	0.0757	A	51.4725	0.9992	0.0381
B	46.0721	0.9975	0.0709	B	52.5961	0.9994	0.0336
C	45.6227	0.9973	0.0763	C	52.1635	0.9992	0.0355
D	47.0168	0.9978	0.0652	D	53.6431	0.9995	0.0303
E	46.4007	0.9975	0.0699	E	51.8801	0.9990	0.0371
F	47.9479	0.9981	0.0587	F	54.8945	0.9996	0.0260
G	47.7852	0.9984	0.0590	G	56.2741	0.9997	0.0219
Photon budget: 5×10^4				**Photon budget:** 10^6			
A	47.1304	0.9984	0.0621	A	51.8006	0.9992	0.0368
B	47.9887	0.9985	0.0563	B	52.9551	0.9994	0.0324
C	46.9508	0.9984	0.0634	C	52.6209	0.9992	0.0338
D	48.4895	0.9987	0.0537	D	53.9767	0.9995	0.0292
E	48.2293	0.9984	0.0551	E	52.0197	0.9990	0.0365
F	48.7697	0.9989	0.0515	F	55.4291	0.9996	0.0245
G	51.4834	0.9992	0.0381	G	56.7556	0.9997	0.0206

E.5 Fidelity-Runtime Evaluation

Same-GPU Monte Carlo timing. The primary runtime comparison uses an NVIDIA GeForce RTX 5080 for both MC simulation and network inference. MCmatlab runs under MATLAB R2025b with native GPU acceleration enabled (`model.MC.useGPU=true`). The computational grid is

64^3, the model width is 0.64 cm, and the fixed irregular-lesion geometry and optical properties match the representative CPU timing configuration. Only `runMonteCarlo` is timed using `tic/toc`, with the stopping criterion set by `nPhotonsRequested`; geometry construction, visualization, and disk I/O are excluded. Each budget from 10^2 to 10^6, and the direct 10^8-photon baseline, has ten timed repetitions. The direct 10^8-photon baseline takes 6.5482 ± 0.0563 s; its ten recorded runs each launch exactly 10^8 photons. This mean provides the reference runtime in the speedup ratio. The 10^9-photon simulation is used only as an independent fidelity reference, not as the timing baseline.

Reconstruction timing and scope. Inference uses one complete 64^3 volume already resident on the RTX 5080, batch size 1, ten warm-up passes, and ten timed forward passes, with CUDA synchronization immediately before and after each pass. The unrounded mean and SD are 0.02251976 s and 0.00052830 s, respectively, reported as 0.0225 ± 0.0005 s. Model loading, disk I/O, CPU-to-GPU transfer, log preprocessing, resampling, and forward/inverse Anscombe transforms are excluded. All methods in Table E7 use the same benchmark protocol; slice-wise methods process all 64 slices, so their time also covers one reconstructed volume.

For input budget N, the reported combined runtime and speedup are

$$T_{\text{sum}}(N) = T_{\text{GPU-MC}}(N) + T_{\text{recon}}, \qquad S(N) = \frac{T_{\text{GPU-MC}}(10^8)}{T_{\text{sum}}(N)}.$$

The direct 10^8 baseline has no reconstruction step. This is the sum of timed simulation and inference components, not a measurement of the complete data-processing pipeline. Under an independence assumption, the SD of the sum is propagated as $s_{\text{sum}} = \sqrt{s_{\text{MC}}^2 + s_{\text{recon}}^2}$; it is not the standard error of the mean or an SD measured from jointly timed pipeline runs. Computations use unrounded means and SDs before display rounding. Timing repetitions characterize one fixed configuration, whereas fidelity statistics average over 81 held-out configurations.

Sampling efficiency versus runtime acceleration. Sampling efficiency compares the input photon histories needed by tested operating points under a common evaluation reference. Section 4.2 compares 10^3-photon reconstruction with raw 10^6-photon MC against the same matched 10^8 references: PSNR/SSIM/RMSE are 42.6318/0.9963/0.1078 versus 38.6475/0.9597/0.1643, respectively. Thus the reconstructed point uses 1000 times fewer input photons while improving all three test-set mean log-domain metrics. This statement concerns aggregate metrics, not dominance on every individual test volume or a measured minimum budget for a quality target. At these two operating points, the same-GPU times are 0.7726 s (10^3 MC plus inference) and 0.7870 s (raw 10^6 MC, with no reconstruction), a runtime ratio of approximately 1.02. Timing characterizes the fixed representative geometry, not the average runtime across the 81 test geometries.

The $8.09\times$ acceleration uses a different pair: 10^6 MC plus inference versus direct 10^8 MC, evaluated for fidelity against the common independent 10^9 reference. Their mean PSNRs differ by 0.6708 dB, so this is a budget-specific runtime comparison at a quantified fidelity gap. An equal-quality comparison requires a specified metric or tolerance, a common reference, and timed direct-MC and reconstruction operating points satisfying that criterion. The present data do not establish the direct-MC runtime at 54.8538 dB, and no equal-quality speedup is inferred by interpolating photon counts or applying an assumed MC scaling law. Budget savings refer only to online input histories; costs of training and high-photon training-data generation are not included.

Independent high-photon reference. Separate 10^8- and 10^9-photon MC simulations use identical geometry and optical properties for each of the same 81 test IDs, paired by their stored volume index. They are independently simulated photon histories, not subsamples of a shared realization. The trained network and its reconstructions are unchanged: no retraining, tuning, or checkpoint selection uses the 10^9 references. The same volume-level PSNR implementation evaluates the inverse-Anscombe output in the stored log domain, using the dynamic range of the selected reference as specified in Appendix C.4. Reported PSNRs are means over the 81 volumes. Direct 10^8 MC is also evaluated against the 10^9 reference, yielding 55.5246 dB. This enables a common-reference comparison to reconstructed fields; it does not define a noise-free ceiling, physical-domain accuracy, or equivalence to a 10^9-photon simulation.

Table E7: Reconstruction runtime of the evaluated methods for one complete 64^3 volume.

Method	**Reconstruction mode**	**Reconstruction time (s)**
3D U-Net	Volumetric	0.0088 ± 0.0003
DnCNN-U-Net	Volumetric	0.0251 ± 0.0005
3D FNO	Volumetric	0.0050 ± 0.0001
MC-Fourier	Volumetric	0.0153 ± 0.0025
ReconFormer	Slice-wise	0.3328 ± 0.0073
VEMamba	Volumetric	0.1259 ± 0.0255
DenoMamba	Slice-wise	0.1472 ± 0.0129
Slice-wise counterpart	Slice-wise	0.0186 ± 0.0049
Proposed	Volumetric	0.0225 ± 0.0005

Table E8: Complete same-GPU fidelity-runtime measurements underlying Table 5.

Photon budget	PSNR vs. 10^8	PSNR vs. 10^9	GPU-MC time (s)	MC + inference time (s)	Speedup
10^2	34.0070	34.9975	0.7700 ± 0.0175	0.7926 ± 0.0175	8.26×
10^3	42.6318	43.4434	0.7501 ± 0.0267	0.7726 ± 0.0267	8.48×
10^4	47.7852	48.2478	0.7423 ± 0.0137	0.7649 ± 0.0137	8.56×
5×10^4	51.4834	51.2789	0.7694 ± 0.0231	0.7919 ± 0.0231	8.27×
10^5	53.1098	52.5179	0.7786 ± 0.0117	0.8012 ± 0.0118	8.17×
2×10^5	54.5029	53.4508	0.7832 ± 0.0285	0.8057 ± 0.0285	8.13×
5×10^5	56.2741	54.4979	0.7645 ± 0.0104	0.7870 ± 0.0104	8.32×
10^6	56.7556	**54.8538**	0.7870 ± 0.0189	0.8095 ± 0.0189	8.09×
Direct 10^8 MC	—	55.5246	6.5482 ± 0.0563	6.5482 ± 0.0563	1×

PSNR (dB) is averaged over 81 test volumes. Both timing components use the same RTX 5080; MC timing uses one fixed configuration. Speedup is relative to direct 10^8-photon GPU-MC. The sum excludes transforms, data transfer, and I/O. Runtimes are mean $\pm$ SD over ten repetitions per component; the SD of their sum is propagated under independence. Reconstruction takes 0.0225 ± 0.0005 s. Speedups use unrounded means.

Alternative CPU-MC/GPU-inference workflow. Table E9 preserves the original timing measurements on an AMD Ryzen 9 9950X (16-core, 32-thread) CPU, using all available cores with MCmatlab GPU acceleration disabled under MATLAB R2025b. The fixed irregular-lesion configuration has lesion FWHM 0.5 mm and the same 64^3 grid. Only MC execution is timed; geometry construction, visualization, and disk I/O are excluded. Input budgets 10^2–10^6 have ten repetitions, while the direct 10^8 CPU baseline has five. Reconstruction uses the RTX 5080 protocol above. These data describe the heterogeneous CPU-MC/GPU-inference workflow quoted separately after the same-GPU result in Section 4.6. Speedup divides direct 10^8-photon CPU-MC time by lower-budget CPU-MC plus GPU inference; it is not a same-GPU or equal-quality ratio. The tabulated combined-time SDs equal the CPU MC SDs at the displayed precision; adding the much smaller reconstruction variance does not change those four-decimal values.

Table E9: CPU-MC/GPU-inference workflow timing retained as an alternative benchmark. Runtimes are mean ± SD; PSNR uses the matched 10^8 reference.

Photon budget	**Output PSNR**	MC runtime (s)	MC + inference runtime (s)	Speedup
$\mathbf{10^2}$	34.0070	0.4643±0.0121	0.4868±0.0121	2493×
$\mathbf{10^3}$	42.6318	0.4614±0.0111	0.4839±0.0111	2508×
$\mathbf{10^4}$	47.7852	0.5571±0.0123	0.5797±0.0123	2094×
$\mathbf{5\times10^4}$	51.4834	1.0045±0.0177	1.0270±0.0177	1182×
$\mathbf{10^5}$	53.1098	1.5579±0.0228	1.5804±0.0228	768×
$\mathbf{2\times10^5}$	54.5029	2.6605±0.0365	2.6830±0.0365	452×
$\mathbf{5\times10^5}$	56.2741	5.9714±0.0602	5.9939±0.0602	202×
$\mathbf{10^6}$	56.7556	11.4705±0.0573	11.4930±0.0573	106×
$\mathbf{10^8}$	—	1213.6854±65.7057	1213.6854±65.7057	1×

Reconstruction takes 0.0225 ± 0.0005 s at all reconstructed budgets; no neural reconstruction is applied at 10^8.

Speedup is relative to direct 10^8-photon CPU-MC and describes a CPU-MC/GPU-inference workflow. At 10^6 photons, $1213.6854/11.4930 \approx 105.60$, reported as $106\times$. The sum excludes preprocessing, transforms, transfer, and I/O.

E.6 Configuration-Paired Reconstruction Comparisons

We evaluate paired per-configuration PSNR differences against 3D FNO and the strongest competing architecture by observed mean PSNR at each budget. The macro-average comparator is selected separately by eight-budget mean PSNR. For configuration i, the macro difference is the mean of its eight budget-specific differences. We resample 81 configuration IDs with replacement 10,000 times (seed 0), preserving each configuration's budget group, and report percentile 95% intervals. A win means a strictly positive difference, with macro wins evaluated after within-configuration averaging.

Index and photon-budget files were checked to agree across all eight methods, with 648 rows ordered as eight blocks of 81 configurations. Generated arrays follow the corresponding test-loader order. The paired analysis recomputes per-volume log-domain PSNR in float64 from saved arrays; small last-decimal differences from the original metric exports do not change the reported comparisons. Intervals are conditional on the trained models and selected comparators, with no adjustment for comparator selection or multiple comparisons. They do not quantify training-seed variation.

Table E10: Configuration-paired PSNR differences (Proposed minus comparator). Intervals use configuration bootstrap resampling.

Budget	Comparator	ΔPSNR (dB)	95% interval	Wins / 81
10^2	3D FNO*	-0.1771	[-0.8285, +0.4668]	42 / 81
10^3	3D FNO	+5.3714	[+4.2524, +6.4644]	70 / 81
10^3	VEMamba*	+2.8180	[+2.3265, +3.3212]	74 / 81
10^4	3D FNO	+5.9059	[+4.9139, +6.9019]	73 / 81
10^4	VEMamba*	+5.1062	[+4.7408, +5.4758]	81 / 81
5×10^4	3D FNO*	+4.6582	[+3.7567, +5.5608]	69 / 81
10^5	3D FNO	+4.6899	[+3.7457, +5.6338]	67 / 81
10^5	ReconFormer*	+4.4743	[+4.1742, +4.7703]	81 / 81
2×10^5	3D FNO	+5.2782	[+4.3447, +6.2436]	74 / 81
2×10^5	ReconFormer*	+4.1682	[+3.9163, +4.4150]	81 / 81
5×10^5	3D FNO	+6.8718	[+5.8425, +7.9180]	71 / 81
5×10^5	ReconFormer*	+4.3299	[+4.0377, +4.6196]	81 / 81
10^6	3D FNO	+7.2601	[+6.1131, +8.4283]	72 / 81
10^6	ReconFormer*	+4.0576	[+3.8247, +4.2888]	81 / 81
Macro-average	3D FNO	+4.9823	[+4.2688, +5.7216]	80 / 81
Macro-average	ReconFormer*	+4.8183	[+4.5663, +5.0717]	81 / 81

*Strongest observed comparator for that budget or macro-average. When 3D FNO is also strongest, one row serves both comparisons. Macro wins concern configuration-level eight-budget means, not all 648 individual inputs.

Figure E2 summarizes the paired comparisons. The macro gain is 4.9823 dB over 3D FNO and 4.8183 dB over ReconFormer, with positive differences in 80 and 81 configurations, respectively. All seven budget-specific intervals from 10^3 to 10^6 exclude zero against the strongest observed comparator. At 10^2, the interval crosses zero; this establishes neither superiority nor equivalence.

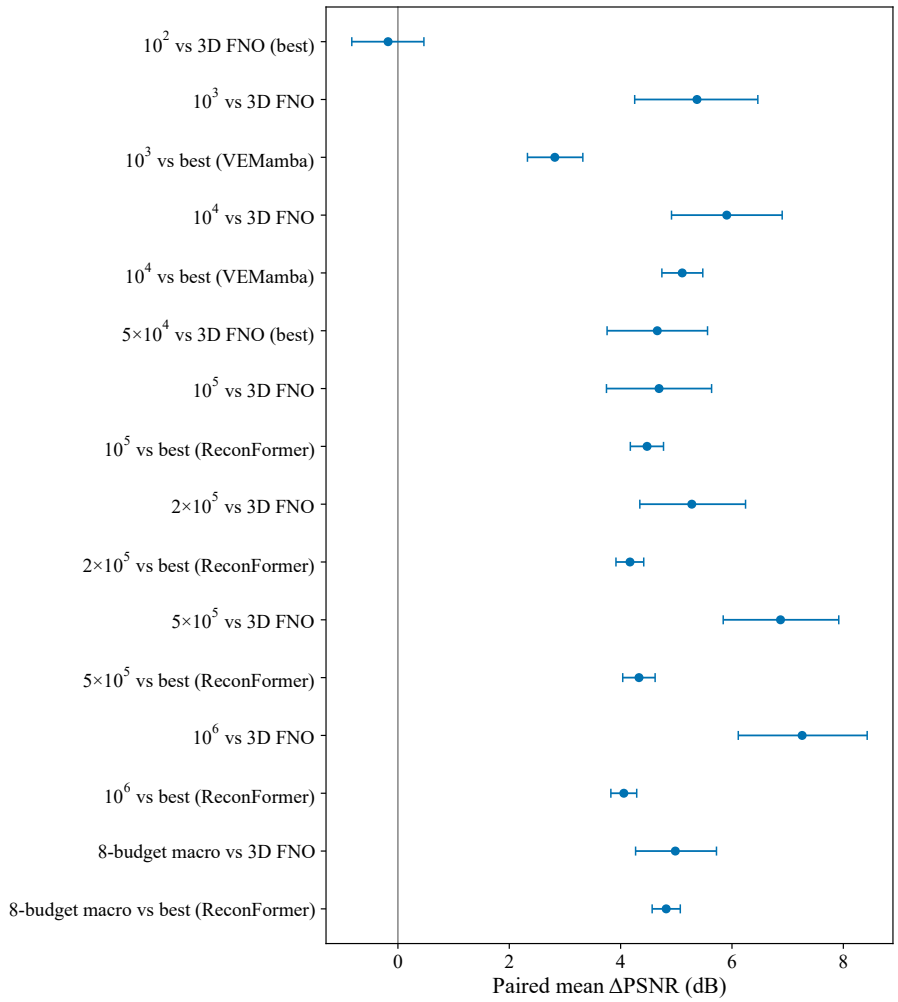

Figure E2: Paired mean PSNR differences and configuration-bootstrap 95% intervals for 81 test configurations. The vertical line marks zero difference; "best" denotes the strongest observed comparator at the stated budget or macro-average. Macro differences preserve the eight-budget group for each configuration.

E.7 Inverse-Transformed Absorption-Field Evaluation

Table E11 evaluates Raw MC, 3D FNO, MC-Fourier, and Proposed against matched 10^8-photon references using the inverse-transformed field metrics in Appendix C.4. All arrays share the same configuration order, stored log transform, and 64^3 grid; no additional test-time normalization or clipping is introduced by this evaluation. The comparison complements log-domain PSNR, SSIM, and RMSE; Figure E3 illustrates the associated depth profiles.

Table E11: Integrated absorption and depth-profile relative errors (%) on the inverse-transformed evaluation grid. Lower is better; bold marks the lowest error within each budget and metric.

	Raw MC		3D FNO		MC-Fourier		Proposed	
Budget	E_{int}	E_{depth}	E_{int}	E_{depth}	E_{int}	E_{depth}	E_{int}	E_{depth}
10^2	46.44	83.90	17.54	35.27	13.76	**29.60**	**13.00**	31.53
10^3	46.10	68.54	17.79	21.73	6.55	10.69	**5.60**	**10.69**
10^4	45.87	67.73	10.14	10.78	5.90	9.33	**3.62**	**5.06**
5×10^4	45.87	67.81	6.05	6.95	3.06	5.46	**1.79**	**2.79**
10^5	45.88	67.86	4.87	5.69	2.63	4.92	**2.14**	**3.21**
2×10^5	45.86	67.83	4.67	5.18	2.58	4.91	**1.92**	**3.06**
5×10^5	45.83	67.84	4.86	5.20	2.66	5.10	**1.93**	**3.03**
10^6	45.88	67.88	4.86	5.09	2.68	5.13	**1.70**	**2.69**
Macro-average	45.97	69.92	8.85	11.99	4.98	9.39	**3.96**	**7.76**

Each entry averages 81 configurations; the macro-average weights all eight budgets equally. All methods use the same matched reference. The field is obtained by inverse log transformation after the common preprocessing and resampling, not by direct evaluation of native-grid MC output.

Proposed has the lowest integrated error at all eight budgets and the lowest depth-profile error at seven. Macro errors decrease by 20.35% and 17.43% relative to MC-Fourier. At 10^6, errors are 1.7037% and 2.6855%. MC-Fourier has lower depth-profile error at 10^2 (29.5957% versus 31.5301%); the 10^3 values are nearly tied (10.6911% versus 10.6861%).

Raw MC retains approximately 46% integrated error at higher budgets despite verified sample alignment. Inverse log transformation emphasizes high-valued voxels; small log-domain errors need not imply small integral errors. This persistent offset is observed under the stated preprocessing and metrics, without establishing its complete cause or an intrinsic native-solver bias.

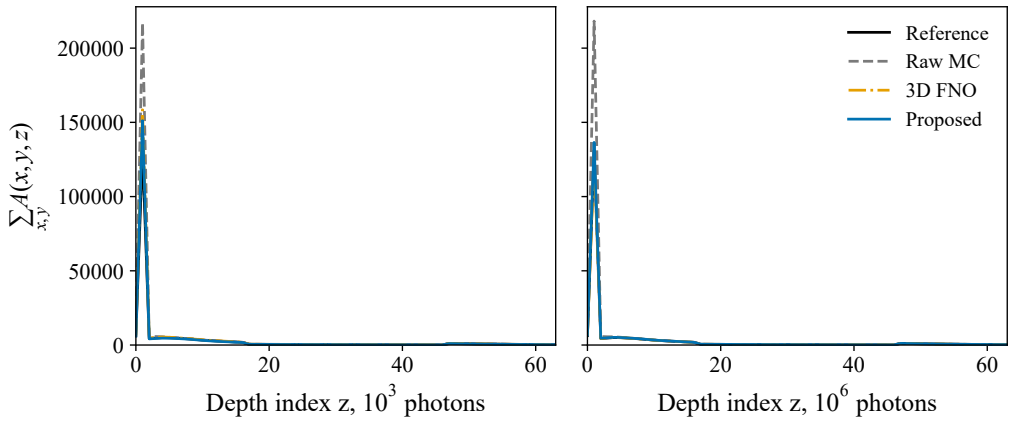

Figure E3: Inverse-transformed depth profiles for configuration 36 at 10^3 (left) and 10^6 (right) photons. The reference is matched 10^8-photon MC. This configuration is the median of the 81 tests ranked by Proposed's mean depth-profile error across eight budgets. Curves illustrate one configuration; Table E11 includes MC-Fourier and aggregates all tests.

E.8 Capacity-Matched Spatial-Domain Mamba Control

We replace the five FACA branches of SMDG + FACA (variant F in Table 4) with a spatial-domain Mamba branch, retaining SMDG and the high-resolution DSA blocks and disabling PBAM. After group normalization and a $1 \times 1 \times 1$ projection from C to $2C$ channels, the branch flattens the spatial grid into a sequence, applies Mamba-3, projects back to C channels, and applies two $3 \times 3 \times 3$ convolutions with an intermediate GELU. Hidden convolution widths are 29, 57, 57, 57, and 29 at Enc3, the three bottlenecks, and Dec3. Mamba uses $d_{\text{state}} = 16$, expansion 2, head dimension 64, and chunk size 64, matching these settings in FACA.

The run uses the existing ablation training and evaluation pipeline on the common configuration split, with no pretrained initialization. The recorded validation-selected best epoch is 165, with validation loss 0.00854369; training stops after 185 epochs. There is one training run for the new control. Table E12 compares its saved test results with the existing convolutional and FACA controls under the same log-domain metric protocol.

Table E12: Spatial-domain Mamba versus convolutional and FACA controls. PSNR (dB) uses matched 10^8-photon references; Δ is FACA minus Spatial Mamba. All three models retain SMDG and disable PBAM.

Budget	SMDG + Conv	SMDG + Spatial Mamba	SMDG + FACA	ΔPSNR
10^2	33.8300	33.1415	**34.4066**	+1.2651
10^3	40.4793	41.7864	**41.9079**	+0.1215
10^4	46.4007	47.8279	**47.9479**	+0.1200
5×10^4	48.2293	48.7282	**48.7697**	+0.0415
10^5	50.0112	50.7637	**50.9433**	+0.1796
2×10^5	51.1917	52.6986	**53.1281**	+0.4295
5×10^5	51.8801	54.3315	**54.8945**	+0.5630
10^6	52.0197	54.7935	**55.4291**	+0.6356
Macro-average	46.7552	48.0089	**48.4284**	+0.4195

The recorded instantiated models contain 4,122,230 parameters for Spatial Mamba and 4,130,193 for the FACA reference, a 0.193% difference. This counting includes a retained 8×64 budget embedding in the no-PBAM model; excluding these 512 entries gives 4,121,718 and 4,129,681, respectively, consistent with the effective-model convention for variant F in the main tables. The convolutional control has 4,126,891 parameters under that convention.

Spatial Mamba improves macro PSNR over the convolutional control by 1.2537 dB; FACA improves it by another 0.4195 dB. FACA has a higher observed mean at every budget, with differences from 0.0415 to 1.2651 dB. The replacement also changes projections and the local mixing/modulation structure, so the experiment tests the FACA branch against a capacity-matched spatial alternative, not Fourier transformation in isolation. It provides neither repeated-training uncertainty nor a compute-matched comparison.

E.9 Complex-Feature Diagnostics for FACA

For the trained Proposed model, we record $Z_b = \mathcal{F}^{-1}(H_b)$ before real projection at all five FACA stages. Norms cover channels and spatial positions for each test input. With $\epsilon = 10^{-12}$,

$$\rho_{\text{imag}} = \frac{\|\operatorname{Im} Z_b\|_2}{\|\operatorname{Re} Z_b\|_2 + \epsilon}, \qquad r_H = \frac{\|H_b(k) - \overline{H_b(-k)}\|_2}{\|H_b\|_2 + \epsilon}.$$

The reconstruction continues through the same real projection and subsequent convolutions. Table E13 summarizes 648 observations per stage; Overall pools all 3,240 stage-input observations. Figure E4 averages the norm ratio over 81 configurations within each budget and stage.

Table E13: FACA imaginary-to-real L_2 norm ratios before real projection and relative Hermitian residuals. These are feature diagnostics, not reconstruction errors or physical energy fractions.

Stage	Mean ρ_{imag}	Median	95th percentile	Maximum	Mean r_H
Enc3	0.8906	0.8918	0.9966	1.0843	1.3272
Bottleneck 1	0.9726	0.9089	1.2534	1.3098	1.3826
Bottleneck 2	0.9329	0.9442	0.9963	1.0401	1.3627
Bottleneck 3	0.9245	0.9249	0.9506	0.9943	1.3575
Dec3	0.7155	0.7161	0.7407	0.7705	1.1636
Overall	0.8872	0.9096	1.0848	1.3098	1.3187

Mean stage ratios range from 0.7155 to 0.9726, with an overall mean of 0.8872. Imaginary features are therefore not numerical roundoff, and the learned spectrum is not approximately Hermitian under this diagnostic. The real projection has the exact Hermitian-projection interpretation given in Appendix A.5; the results do not show that discarding the imaginary component is lossless or optimal.

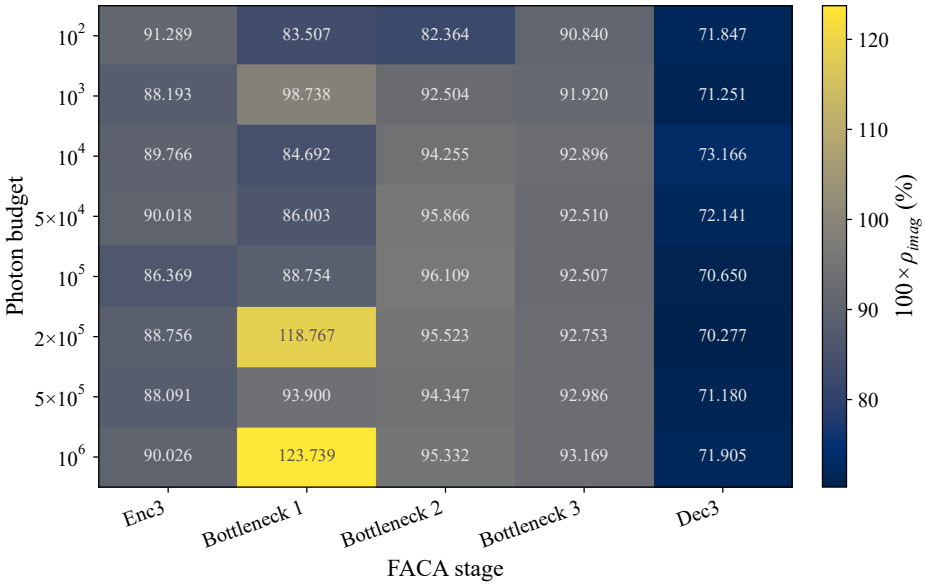

Figure E4: Mean imaginary-to-real feature norm ratio by budget and FACA stage. The heatmap displays $100\rho_{\text{imag}}$: 100% means equal imaginary and real L_2 norms, not that imaginary features contain all of the total energy. The squared-norm ratio would be ρ^2_{imag}.

F Held-Out Physical-Regime Generalization

F.1 Splits and Training Protocol

These experiments reuse the 405 selected configurations and existing MC volumes without new simulation. They define two separate partitions, leaving the main 243/81/81 split unchanged. Within each experiment, training, validation, ID test, and held-out test configurations are mutually disjoint; all eight input budgets of a configuration remain together. The two experiments need not have disjoint configurations from each other. Table F1 gives configuration and input-volume counts.

For lesion absorption, training and validation contain only 0.5 and 1.0 cm^{-1}, ID testing uses 1.0 cm^{-1}, and all 142 configurations at 1.5 cm^{-1} form the held-out set. For lesion scale, training and validation contain FWHM 0.5, 0.8, and 1.2 mm, ID testing uses 1.2 mm, and all 101 configurations at 1.6 mm form the held-out set. FWHM is $2\sqrt{2\ln 2}\,\sigma$, where σ is the Gaussian morphology-scale parameter. This tests larger lesion scales, not a new shape family. Invasion depth varies independently over ten values from 0.113 to 0.131 cm in steps of 0.002 cm. Layer structure and lesion placement follow the original generator.

Table F1: Physical-regime splits. Each cell lists configurations / input volumes; each configuration contributes eight budgets sharing one reference.

Held-out variable	Train	Validation	ID test	Held-out test
Lesion absorption	185 / 1480	39 / 312	39 / 312	142 / 1136
Lesion scale	212 / 1696	46 / 368	46 / 368	101 / 808

ID and held-out samples are not paired. To improve comparability, 1,000 candidate ID subsets are sampled from the nearest seen value, and the subset minimizing the sum of marginal-distribution L1 distances for the other five parameters is selected. This uses metadata only, never reconstruction results. Selected seeds are 143 and 74, with distances 0.574 and 0.833 for absorption and scale, respectively. Remaining eligible configurations are randomly divided into training and validation. Matching is approximate and marginal, rather than a control of every joint parameter combination. Tables F5 and F6 report all six parameter distributions.

Each method is trained from random initialization separately for each split, yielding six models. Main-experiment checkpoints are not reused because their training sets include the held-out values. The experiment runner invokes each method's training implementation with the new data loaders and result directory. Training runs for at most 200 epochs, with early-stopping patience 20. The checkpoint with the lowest validation objective is used for both ID and held-out evaluation without adaptation or reselection. Best epochs are 194/190 for 3D FNO, 130/121 for MC-Fourier, and 152/110 for the proposed model (absorption/scale). Runs use the same software environment. Split seeds are fixed, but network initialization and minibatch shuffling seeds were not fixed; there is one training run per method and split. Different training and ID sets explain why ID PSNR differs between the two experiments.

F.2 Evaluation and Uncertainty

Each input has an independently simulated, configuration-matched 10^8-photon reference. All eight budgets share the same reference for a configuration. Evaluation uses complete 64^3 volumes, including background, in the stored log domain $u = \ln(1 + e^8 a)$. For Anscombe-transformed outputs, the inverse transform is applied before exported arrays are evaluated; it restores u, not physical absorption a. The 3D FNO output is evaluated on the same stored log scale. The evaluator applies no further Anscombe or inverse-log transform. PSNR and SSIM use each reference volume's maximum minus minimum as dynamic range, following Appendix C.4.

Metrics are averaged across configurations within each budget, then equally across the eight budgets. PSNR drop is the unrounded ID mean minus the unrounded held-out mean. For uncertainty, each configuration's PSNR is first averaged over its eight budgets. SD describes variation between configurations. Percentile 95% confidence intervals use 10,000 configuration-level bootstrap resamples; ID and held-out sets are independently resampled for drop intervals. Budgets are not treated as inde-

pendent configurations. These intervals quantify test-configuration sampling uncertainty conditional on a fitted model and exclude training randomness. They are not paired intervals for differences between methods.

Table F2: PSNR mean ± configuration SD and 95% bootstrap intervals; drop intervals use independent ID and held-out resampling. Values use eight-budget configuration averages.

Method	ID PSNR (dB)	Held-out PSNR (dB)	Drop (dB)
Lesion absorption			
3D FNO	42.522 ± 4.65 [41.08, 43.96]	41.797 ± 4.71 [41.02, 42.54]	0.725 [-0.89, 2.34]
MC-Fourier	42.821 ± 1.84 [42.24, 43.39]	42.584 ± 2.09 [42.23, 42.92]	0.237 [-0.42, 0.90]
Proposed	48.067 ± 2.48 [47.30, 48.83]	47.927 ± 2.43 [47.52, 48.32]	0.140 [-0.75, 1.00]
Lesion scale			
3D FNO	41.741 ± 4.16 [40.57, 42.90]	39.192 ± 3.29 [38.55, 39.83]	2.549 [1.24, 3.89]
MC-Fourier	42.591 ± 1.64 [42.12, 43.05]	40.750 ± 1.95 [40.37, 41.13]	1.842 [1.23, 2.44]
Proposed	45.049 ± 1.47 [44.62, 45.47]	43.056 ± 1.74 [42.72, 43.39]	1.994 [1.44, 2.53]

Table F3: SSIM and RMSE macro-averaged equally across eight photon budgets.

Method	ID SSIM	Held-out SSIM	ID RMSE	Held-out RMSE
Lesion absorption				
3D FNO	0.9963	0.9962	0.1481	0.1544
MC-Fourier	0.9928	0.9925	0.1294	0.1299
Proposed	0.9964	0.9963	0.0948	0.0941
Lesion scale				
3D FNO	0.9960	0.9926	0.1561	0.1961
MC-Fourier	0.9928	0.9893	0.1365	0.1637
Proposed	0.9958	0.9932	0.1104	0.1348

The proposed model has the highest held-out PSNR and SSIM and lowest held-out RMSE among the three methods in both experiments. All absorption-drop intervals include zero, whereas all lesion-scale-drop intervals are positive. Neither observation establishes a difference between methods' drops. The proposed model's slightly lower absorption held-out RMSE alongside a positive PSNR drop is possible because the sets are unpaired, reference dynamic ranges vary, and logarithmic PSNR and RMSE are averaged separately. These results support retained reconstruction advantages under the tested extrapolations, with greater degradation in the lesion-scale experiment; the shifts and training sets differ, so this does not establish that scale changes are universally harder.

F.3 Results Across Photon Budgets

Table F4: Per-budget log-domain PSNR (dB) against matched 10^8-photon references. Raw MC is the unreconstructed input. Macro-averages use unrounded values.

Method	Set	10^2	10^3	10^4	5×10^4	10^5	2×10^5	5×10^5	10^6
Lesion absorption									
3D FNO	ID	31.98	35.29	40.65	44.68	46.28	46.82	47.20	47.28
3D FNO	Held-out	32.06	35.22	40.28	44.02	45.07	45.71	46.01	46.00
MC-Fourier	ID	31.76	37.80	41.23	44.09	45.59	46.65	47.58	47.86
MC-Fourier	Held-out	32.07	37.31	41.05	43.96	45.31	46.34	47.16	47.45
Proposed	ID	32.32	40.31	46.35	50.17	51.83	53.39	54.65	55.51
Proposed	Held-out	32.46	39.57	46.11	50.11	51.72	53.36	54.60	55.48
Raw MC	ID	7.42	8.67	12.97	19.69	24.04	28.97	35.11	38.83
Raw MC	Held-out	7.43	8.69	12.97	19.70	24.05	28.98	35.07	38.81
Lesion scale									
3D FNO	ID	31.39	34.55	40.16	44.27	45.30	46.04	46.14	46.07
3D FNO	Held-out	28.21	31.21	37.44	41.78	42.96	43.66	44.08	44.19
MC-Fourier	ID	30.67	37.29	41.52	44.44	45.65	46.48	47.24	47.45
MC-Fourier	Held-out	27.52	36.11	40.00	43.06	44.03	44.69	45.22	45.38
Proposed	ID	31.65	39.42	43.98	47.16	48.17	49.33	50.13	50.55
Proposed	Held-out	28.64	37.57	42.02	45.36	46.54	47.51	48.20	48.61
Raw MC	ID	7.35	8.62	12.96	19.76	24.16	29.03	35.13	38.86
Raw MC	Held-out	6.52	7.90	12.56	19.83	24.43	29.52	35.50	39.13

The proposed model has the highest held-out PSNR at every evaluated budget in both experiments. In the absorption experiment, its held-out PSNR approaches ID PSNR at high budgets; in the lesion-scale experiment, a gap persists across all budgets. These comparisons describe the fitted models on their respective test sets and do not isolate a causal effect of the held-out variable.

F.4 Parameter Distributions

Counts below are configuration counts, not counts of individual photon-budget volumes. Held-out values are absent from training and validation in their respective experiments; the other experiment may include those values because it holds out a different variable.

Table F5: Configuration distributions for the lesion absorption experiment.

Parameter	Value	Train	Validation	ID test	Held-out
Melanin absorption (cm^{-1})	8.0	67	19	13	50
	14.0	66	11	16	47
	20.0	52	9	10	45
Dermal absorption (cm^{-1})	0.2	57	14	16	62
	0.4	57	14	11	41
	0.6	71	11	12	39
Lesion absorption (cm^{-1})	0.5	111	23	0	0
	1.0	74	16	39	0
	1.5	0	0	0	142
Lesion reduced scattering (cm^{-1})	10.0	70	12	12	43
	14.0	54	13	15	56
	18.0	61	14	12	43
Lesion scale FWHM (mm)	0.5	55	10	11	37
	0.8	40	10	8	32
	1.2	48	9	10	34
	1.6	42	10	10	39
Invasion depth (cm)	0.113	18	6	6	17
	0.115	15	2	4	10
	0.117	17	3	3	13
	0.119	21	1	4	12
	0.121	21	5	5	16
	0.123	18	5	2	16
	0.125	28	3	4	15
	0.127	9	4	3	10
	0.129	19	4	6	19
	0.131	19	6	2	14

Table F6: Configuration distributions for the lesion scale experiment.

Parameter	Value	Train	Validation	ID test	Held-out
Melanin absorption (cm^{-1})	8.0	82	13	17	37
	14.0	73	20	15	32
	20.0	57	13	14	32
Dermal absorption (cm^{-1})	0.2	82	18	14	35
	0.4	63	12	15	33
	0.6	67	16	17	33
Lesion absorption (cm^{-1})	0.5	72	13	17	32
	1.0	69	18	12	30
	1.5	71	15	17	39
Lesion reduced scattering (cm^{-1})	10.0	68	11	16	42
	14.0	71	23	16	28
	18.0	73	12	14	31
Lesion scale FWHM (mm)	0.5	92	21	0	0
	0.8	75	15	0	0
	1.2	45	10	46	0
	1.6	0	0	0	101
Invasion depth (cm)	0.113	30	8	5	4
	0.115	17	2	5	7
	0.117	18	4	5	9
	0.119	24	3	2	9
	0.121	24	7	4	12
	0.123	18	2	4	17
	0.125	26	5	6	13
	0.127	16	2	0	8
	0.129	20	7	9	12
	0.131	19	6	6	10